DOCUMENT RESUME

ED 314 432 TM 014 213

AUTHOR	Pfleiderer, Joanne, Ed.
TITLE	New Directions for Teacher Assessment. Proceedings of the ETS Invitational Conference (49th, New York, New York, October 29, 1988).
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ISBN-0-88685-089-4
PUB DATE	89
NOTE	121p.
AVAILABLE FROM	Educational Testing Service, Invitational Conference Proceedings, Princeton, NJ 08541 ($12.95 if prepaid, $15.95 if not prepaid).
PUB TYPE	Collected Works - Conference Proceedings (021) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Beginning Teachers; *Educational Trends; Elementary Secondary Education; Higher Education; *National Programs; Student Teacher Supervisors; Teacher Attitudes; Teacher Certification; Teacher Education; *Teacher Evaluation
IDENTIFIERS	Rochester Method

ABSTRACT

Proceedings of a conference on teacher assessment are presented. A speech accompanying presentations of the 1988 Educational Testing Service Award for Distinguished Service to Measurement and a list of winners of that annual award from 1970 to 1988 are included. The nine papers provided include: (1) "Strengthening the Teaching Profession through National Certification" (James A. Kelly); (2) "The Paradox of Teacher Assessment" (Lee S. Shulman); (3) "A New Generation of Tests for Licensing Beginning Teachers" (Carol Anne Dwyer); (4) "Implications of Studies of Expertise in Pedagogy for Teacher Education and Evaluation" (David C. Berliner); (5) "A Classroom Teacher's View of the Assessment of Teaching" (Claire L. Pelton); (6) "New Directions for the Career of Teaching--The Rochester Experiment" (Adam Urbanski); (7) "Restructuring Teacher Education" (Judith E. Lanier); (8) "The Case for a Supervised Teaching Internship" (Linda Darling-Hammond); and (9) "Will Improving Teacher Assessment Improve the Education of Children?" (P. Michael Timpane). (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NORMA J. FINNERAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED314432

New Directions for Teacher Assessment

Educational Testing Service

TM014213

BEST COPY AVAILABLE

INVITATIONAL CONFERENCE PROCEEDINGS

New Directions for Teacher Assessment

Proceedings of the 1988 ETS Invitational Conference

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541

The forty-ninth ETS Invitational Conference, sponsored by Educational Testing Service, was held at The Plaza, New York City, on October 29, 1988.

Presiding: Gregory R. Anrig
President
Educational Testing Service

Conference Coordinator: Margaret B. Lamb

Proceedings Editor: Joanne Pfleiderer

Production: Roseann Spano

Educational Testing Service is an equal opportunity/affirmative action employer.

Copyright © 1989 by Educational Testing Service. All rights reserved.

Educational Testing Service, ETS, and ⓔ are registered trademarks of Educational Testing Service.

Library of Congress Catalog Number: ISBN 0-88685-089-4

Contents

v Presentation to Donald T. Campbell of the 1988 ETS Award for Distinguished Service to Measurement

vii ETS Award for Distinguished Service to Measurement, Recipients, 1970–1988

1 Strengthening the Teaching Profession Through National Certification
James A. Kelly

13 The Paradox of Teacher Assessment
Lee S. Shulman

29 A New Generation of Tests for Licensing Beginning Teachers
Carol Anne Dwyer

39 Implications of Studies of Expertise in Pedagogy for Teacher Education and Evaluation
David C. Berliner

69 A Classroom Teacher's View of the Assessment of Teaching
Claire L. Pelton

77 New Directions for the Career of Teaching — The Rochester Experiment
Adam Urbanski

85 Restructuring Teacher Education
Judith E. Lanier

99 The Case for a Supervised Teaching Internship
Linda Darling-Hammond

115 Will Improving Teacher Assessment Improve the Education of Children?
P. Michael Timpane

The 1988 ETS Award for Distinguished Service to Measurement

Presented to
DONALD T. CAMPBELL

Donald Campbell's many distinguished contributions to psychology, sociology, cultural anthropology, and philosophy of science have won him numerous prestigious laurels as well as widespread renown as the *compleat* social scientist.

In further tribute, he is being honored here for his methodological work over the past four decades. During this period, Professor Campbell has persistently alerted the measurement field to a variety of threats to the validity of test interpretation. He has been particularly concerned with the role of method variance in diluting or undermining score meaning. Method variance includes all systematic effects associated with a particular measurement procedure that are extraneous to the attribute or process being measured. In his continuing efforts to cope with the threats of method variance, Donald Campbell has championed such innovations as the indirect assessment of social attitudes and personality, unobtrusive and nonreactive measurement in social science research as well as in program evaluation, and test validation by means of multitrait-multimethod designs. In the process, he has effectively galvanized the measurement field around new principles of validation — namely, the need for both convergent and discriminant evidence corroborating score meaning and discounting plausible rival interpretations.

With equal vigor and persistence, Professor Campbell has also repeatedly probed diverse threats to the tenability and generalizability of experimental inferences in field settings. By formalizing concepts of quasiexperimentation in rigorous and practically useful ways, he helped elevate program evaluation and policy studies to the level of scientific inference. These efforts thereby give credence to the promise of an experimenting society in which reforms or social interventions are conceptualized as experiments to be systematically evaluated.

For his clarifying refinements of construct validity theory, for his powerful formulations of quasiexperimental techniques and of convergent and discriminant validation, and for his insistence that test validation and program evaluation can and should be viewed as scientific inquiry, ETS is pleased to present its 1988 Award for Distinguished Service to Measurement to Donald T. Campbell.

ETS Award for Distinguished Service to Measurement Recipients, 1970 – 1988

1970 E.F. Lindquist

1971 Lee J. Cronbach

1972 Robert L. Thorndike

1973 Oscar K. Buros

1974 J.P. Guilford

1975 Harold Gulliksen

1976 Ralph W. Tyler

1977 Anne Anastasi

1978 John C. Flanagan

1979 Robert L. Ebel

1980 John B. Carroll

1981 Ledyard R Tucker

1982 Raymond B. Cattell

1983 Frederic M. Lord

1984 Henry Chauncey (special award)
Louis Guttman

1985 Paul Horst

1986 Frederic Kuder

1987 Karl G. Jöreskog

1988 Donald T. Campbell

Strengthening the Teaching Profession Through National Certification

JAMES A. KELLY
National Board for Professional Teaching Standards

A voluntary national certification system for teachers is being developed by a new organization, the National Board for Professional Teaching Standards. The mission of the National Board is to establish high and rigorous standards for what teachers should know and be able to do, to establish a system for Board certification, to develop a state-of-the-art assessment system to ascertain who meets the standards for Board certification, and to encourage adoption of related education reforms intended to improve student learning. Board certification is at the heart of efforts to strengthen the teaching profession. Complementing other education reforms, Board certification can lead to markedly improved student learning. It can also provide both recognition of excellent teachers — and there are many — and a standard toward which all teachers can aspire.

In my comments today, I will review the pressing social need to strengthen the teaching profession, describe the National Board for Professional Teaching Standards, suggest standards for good teaching that should characterize the profession, discuss criteria for an effective assessment based on those standards, and offer an early glimpse of the impact Board certification and assessment can be expected to have on the teaching profession in the next few decades.

Why It Is Important to Strengthen the Profession

Dismal assessments of the state of American education have been all too common in the past decade. However accurate the findings or inflammatory the rhetoric, they leave us pondering some hard facts about the challenges facing American schools in an increasingly complex global society.

No longer can we rest smugly secure in the knowledge that the U.S. economy leads the world. We face instead an economy in transition, an economy challenged on many fronts by international competition. It is said that schools are failing to prepare the next generation of workers for the new competitive environment.

The traditional social mission of American schools — to build a cohesive society from a plethora of diverse voices — has never been more relevant. We are a nation growing more diverse every day. But the traditional structures that help socialize children have also undergone rapid changes. The traditional family may never look quite the same again. In many communities, churches no longer play the strong role they once did. Precious few unifying elements — save perhaps pop culture, sports, and the media — bind us together. The common school must be the bedrock of a cohesive society.

The schools, we are told, are trying to do too much and succeeding in doing too litt'e, achieving neither excellence nor equity. At a time when we have never demanded so much of citizens, we are told that our 17-year-olds do not know enough about civics, literature, math, science, geography, and worse, know even less of history and its lessons.

Yet we ask even more of our schools. The very individualism that may be our economic lifeblood must be balanced by a sense of civic duty to country and personal responsibility to one another. Schools should be nurturing that commitment to the public good.

More than ever, then, schools need to achieve what Dewey proclaimed as their mission earlier in this century — the maximum development of each individual. Functioning in a democratic society today, a society framed by a broader global community, students must become competent in the use of technology for learning and communication, understand foreign languages and cultures, and possess far more than basic math and science skills. Students must be able to learn how to think for a living and must possess skills that permit lifelong learning and adaptation to the challenges of the next century.

Faced with this agenda, the nation has never needed excellent teachers more, but new problems have emerged in the teacher work force. The profession's present salary structure, the teacher's role in the classroom, and the low status associated with teaching do little to attract sufficient numbers of qualified new candidates or to encourage the retention of the best and most able practitioners. The problem of teacher supply has also been exacerbated as new windows of opportu-

nity have opened for women and minorities in other fields. And the problem of supply is most critical in the case of minority teachers; simply put, we have too few minority teachers in the classroom, and without such role models we are attracting even fewer to the next generation of teachers.

Worries about the quality of teachers have also been voiced. Though dozens of "accountability measures" have been spawned, many educators, as well as government and business leaders, continue to express doubts concerning the quality of teaching in schools. There are at present no nationally accepted professional teaching standards to which teachers can aspire. Indeed, the authority structure of school organizations often places little faith in the wisdom of teaching practice. Status — at least in terms of money and power over the decisions affecting classrooms — accrues to those who are farthest from the classroom. Little distinction is made between the first-year and twentieth-year teacher.

These factors combine to shape an image of the profession that is often frustrating to those who elect to venture into it. At a time when education is more critical than ever, we must strengthen the profession if we are to meet these challenges to the education system.

The National Board

The National Board for Professional Teaching Standards is a nonprofit organization whose main purpose is to establish high and rigorous standards for what teachers should know and be able to do and to certify teachers who meet the standards. The Board's goal is a strengthened teaching profession that will raise the quality of education available to children in the United States.

Professions such as medicine, architecture, and accounting have national systems of professional certification, but teachers are subject only to state licensing systems that set minimum standards for beginning teachers. Board certification will go far beyond state licensing. Because Board-certified teachers will have met high and rigorous standards, certification will become a public affirmation that a teacher is a true professional who exhibits creativity and imagination in solving complex problems and has command of subject matter and how to convey its central tenets in a variety of ways to all students. Board certification is intended to be a declaration that a teacher can effectively

exercise independent professional judgment on a variety of instructional matters that affect learning. In this way, the Board's requirements will be similar in many respects to the standards other professionals must meet.

While compulsory licensing standards vary from state to state, the Board expects to adopt a single set of standards and assessment practices to be applied uniformly across the country. Such a practice will facilitate state reciprocity agreements and thereby enhance teacher mobility, a valuable benefit at a time when the demand for teachers from state to state is likely to be highly variable.

Board certification is being offered as a voluntary vehicle for teachers to help them help themselves in terms of professional advancement and recognition. Governors, legislators, and state education officials familiar with the Board understand that Board certification is intended to be voluntary. Each state and locality will have to decide for itself in what manner and to what degree it wishes to recognize Board-certified teachers.

While it is unlikely that states will choose to adopt the Board's standard as their licensing standard, there is interest in moving to performance-based licensing examinations and in creating "Board-compatible" licensing systems. For example, the state departments of education in Connecticut and California have established the New Teacher Assessment and Support Consortium to move in just this direction. The plan of Educational Testing Service to design and develop a new National Teacher Examination is an important and welcome step that will surely improve the assessment of persons seeking state licenses to teach.

A 63-person board of directors governs the Board; a majority of the members are actively engaged in teaching elementary- and secondary-school children. The board of directors also includes state and local officials with responsibilities for elementary and secondary education and leaders from the business community and higher education.

Two-thirds of the Board is made up of teaching professionals. Eventually, these teaching professional members will be elected by Board-certified teachers. Until that time, one-third are teacher-union leaders; one-third are teachers who have achieved distinction in subject matter or teaching-specialty fields; and one-third are teachers selected solely on the basis of outstanding records of accomplishment in the classroom.

As a reflection of the many constituencies that are stakeholders in elementary and secondary education, one-third of the Board is made up of public and other educators. This category consists of people whose responsibilities fall in the following fields: governance of education (governors, state legislative leaders, chief state school officers, state and local board of education members); management of education (school principals and superintendents); higher education (university administrators and faculty members); and leaders of the minority and business communities.

The Board has been endorsed by the National Governors' Association, the National Education Association, and the American Federation of Teachers; by editorials in the nation's leading newspapers; and by countless leaders from education, business, and public life. It is widely regarded as having the potential to achieve strategic improvement in the quality of education in the United States.

Standards for Certification

The National Board is expected to adopt a formal statement of teaching standards during the spring of 1989. Although the following comments do not reflect formal Board policy, they are consistent with the direction and substance of Board thinking in late 1988.

Our vision is shaped by an image of teaching that is complex but grounded in knowledge that can be learned and applied by most teachers. We see teachers as informed and principled decision makers, professionals who make thousands of small and large decisions influenced by a host of complicating factors. Like the anthropologist in the field or the scientist in the laboratory, what teachers do is guided by working hypotheses, educated guesses about the nature of what happens. These working hypotheses are constantly tested, revised, and revamped in the light of what is known about teaching and by teachers' own knowledge and experience.

Take, for example, one teacher, one student, and one lesson objective that is presumably unambiguous: teaching the concept of decimals. At any given moment during the lesson, the teacher weighs and acts on a shifting complex of information. What does that child already know? How must this new knowledge be linked to other concepts? What are the special strengths and weaknesses of the learner in abstract and linguistic reasoning? How well does he or she understand

numbers and operations performed on them? What interests does he or she have that might be used in developing stimulating problems? For example, could enthusiasm for baseball be used to transform the concept into a question of batting averages? How is that one learner situated in the classroom or in the smaller group within that classroom? What material — be it a blackboard, an overhead, or some material to manipulate — can be used to demonstrate the concept? How will the knowledge be evaluated? What time of day is it? How long will the student's attention span last? What does the teacher do if the planned attack on the concept simply does not work?

And of course, each working hypothesis or decision is never made for only one learner at a time. The teacher makes decisions for 25 or more children simultaneously, weighing the best answers for the group. Amidst this Babel of competing interests, the teacher is charged with selecting plausible solutions.

Such decision-making skills are not the hallmark of every occupation; rather, they are characteristics of a profession — where practitioners make judgments by drawing on professional skills and knowledge learned from both training and experience. The more one has the chance to hone these decision-making skills, the better they become. Thus professional experience becomes a critical component in developing and evaluating professional competence.

Let us always remember that there is no single right answer to many of the most important questions in teaching.

Accomplished professional teachers are committed to their students and their students' learning. They treat all students equitably and understand the personal and instructional implications of student differences. These teachers are skilled at diagnosing and facilitating learning. Armed with a variety of methods and modes of interaction, they must create, enrich, and on occasion alter the organizational structures in which they work with young people. They find ways to capture and sustain the interest of students; moreover, they are adept at pacing and evaluating their students' learning. They make efficient and effective use of time available to them.

In order to make good decisions, teachers must know the subjects they teach, how to teach those subjects to students, and how their students learn and develop. Teachers organize instructional settings that allow students to develop an appreciation for the interconnections that exist between and among disciplines. They model the modes of inquiry appropriate to the subject. But they are also schooled in the

lore of teaching — how to manage a classroom, how to discipline effectively, how to use curricular materials, how to plan and execute lessons.

Because they are professionals, teachers continually think about and learn from experience. As critical thinkers, teachers are also members of learning communities. They work in collaboration with other teachers in collegial settings to improve school-wide learning. Teachers understand the environments from which their students come and are mindful of their students' many needs.

Professional teaching standards derived from a view of teachers as sophisticated decision makers are most likely to be accepted if they come from an enterprise such as the National Board, which is rooted in the teaching profession itself, yet incorporates the voices of other constituencies and stakeholders in education.

The Assessment Challenge

While the standards I have described represent a sound basis from which to develop an assessment of effective teaching, they also remind us that current tests of teachers are not oriented toward this view. Currently, state licensing examinations assess, almost exclusively through paper-and-pencil technology, a limited range of minimal competencies. But no assessment yet exists in teaching that is based on the ways in which teachers make decisions and exercise professional judgment.

Models for such assessments do exist in other professions, however. When architects or doctors are certified as professionals, the assessment reflects the fact that certified architects or doctors must be able to make good decisions. In the architect's registrant examination, candidates are given a design problem in which they must use their knowledge to design a structure that meets a client's needs. In the patient-management problems of medical licensing examinations, physicians make judgments when considering a specific history and demonstrate that they know which questions to ask in order to make an appropriate diagnosis. In the board-certification examinations for physicians, patient histories are discussed in great detail, and decisions about treatment are examined in light of diagnoses and known outcomes. In the California bar exam, candidates are given a client problem along

with the case history, depositions, and relevant raw material; they then prepare an appropriate response, such as a brief, letter, or appeal.

In each of these cases, certified professional practitioners judge responses on the basis of the context and constraints presented as well as on what is generally considered to be sound professional practice.

No such assessment reflecting knowledge, judgment, and context is in common use in teaching, however.

If we are to take these standards seriously, we must realize that we have set before us an assessment task of enormous proportions, one that will absorb all of us for several years to come. Measuring the kinds of skills and competencies outlined above far outdistances present assessment models in teaching. The technical expertise of the assessment community and the wisdom of practice of teaching professionals must be married in the new assessments. We envision a research and development effort of several years, costing nearly $50 million, to develop such assessments in the diverse fields that characterize the teaching profession. The Board's research plans will be publicly announced so qualified organizations can submit proposals. Research results and findings will be made public so the entire profession can benefit from what is learned.

The National Board is about to begin that research effort. We will be guided, in part, by the standards of the measurement community. Developing assessments that are both valid and reliable is essential. At every stage, we must ask ourselves: Do these assessments measure what they purport to measure? Can the results be relied on with confidence?

Surely an effective assessment must be consistent and fairly administered. Such a benchmark requires that our certification process follow those precepts of assessment development that require both time and care: careful field-testing, thorough training of jurors, and scrupulous examination of results for accuracy and fairness.

But ultimately the validity of this assessment will not be judged in technical terms. The primary criterion for validity of certification within the teaching profession is its recognition as credible by excellent teaching professionals, by other educators, by members of governing boards, and by the public. Certification must align itself with the realities and rigors of practice and represent the best of the profession. The certification developed by the Board will be designed to meet this criterion. Unlike many other teacher tests, this assessment will not be

dictated *to* teachers. Rather, it will be the direct outgrowth of a process that has been guided by leaders of the teaching profession, with the involvement of other educators and representatives of the public. The articulation of these standards by such a diverse group will itself be a noteworthy accomplishment. Assessments based on such standards will be valid in a new way.

Along with reliability and validity, there is, therefore, a third criterion for teacher assessments: impact. We must consider the impact that Board certification is likely to have on the face of American education: on who enters the profession; on the training of teachers; on the context and environment in which they teach; and most importantly, on what students learn.

Impact of National Certification

High and rigorous standards embraced by the teaching profession should bring new rigor and coherence to the preparation of teachers. As excellence is recognized and Board-certified teachers become a reality, we envision a salutary effect on the self-esteem of practitioners receiving such recognition and on the aspirations of young persons contemplating teaching as a career. Board certification will help to retain in teaching a larger proportion of today's best teachers than would otherwise be the case.

The standards for Board certification will have an impact on teaching practice, as teachers think about the standards, prepare for certification, take the assessments, and serve as mentors for other teachers in subsequent assessments. Professional teachers who are Board-certified may be given increased responsibilities for instructional leadership and supervision. Changes should occur in the ways that schools are organized and managed so that today's accomplished teachers can teach more effectively and students can learn more efficiently. As practitioners unite to develop their own standards and administer their own assessment of professional worth, the process of Board certification will increase self-governance within the teaching profession.

In short, we seek a strengthened profession that inspires the confidence of educators as well as of business and government leaders. The ultimate result of Board certification is intended to be an overall improvement in the quality of learning for children in our schools.

National certification is but one of many reforms needed in education. Board certification is not a panacea. There are related concerns that demand attention: the organization of teaching in schools; the supply of teachers, especially minority teachers; and teacher education. Equity is another concern. The Board must insist that the distribution of Board-certified teachers among school districts does not exacerbate existing inequities. Everything possible must be done by policymakers to secure access by rural, poor, and minority students to Board-certified teachers. The Board must also be vigilant and ensure that the standards and resulting assessments are intrinsically fair. All teachers must be treated fairly in the administration of the assessment, in the prerequisites demanded, in the conditions of practice that lead to certification, and in the questions and exercises themselves.

Over two centuries ago, Jefferson's plan for a state system of education included an evaluator whose charge was "to give assurance to the fidelity of the commonwealth" about the quality of classroom teachers. Given that American education is fundamentally a democratic enterprise, demands for accountability are nothing new.

But in recent years, demands for more measures of teacher quality have proliferated. How should the effectiveness of this larger enterprise be measured?

National Board certification should certainly not be evaluated by how quickly Board certification is implemented. While the Board will move forward as fast as is possible, it will also encourage review and discussion of its standards, plans for assessment, and interests regarding educational policy and reform. The organizational structure to support this effort is still being built. The research and development activities that are necessary to secure the very best ways to measure and attest to these standards will soon begin. The Board will rely on those experienced in assessment and committed to good teaching to assist in these activities.

Even given these daunting tasks, Board certification should be a reality within five years. By that time, hundreds of experienced teachers, other educators, and scholars will have participated in its development, and more will have been involved as candidates or jurors in the process of creating the first cohorts of Board-certified teachers.

It may take another five years — bringing us near the year 2000 — before the impact of Board certification on schools is clear. By that time, there will be a strong and committed group of practitioners and scholars who are regularly involved in the development and adminis-

tration of Board certification. And tens of thousands of Board-certified teachers will be teaching in America's schools.

By the end of the first decade of the 21st century, there should be an entire generation of teachers for whom Board certification is a reality. The effects of this cohort will be felt in a strengthened profession and renewed schools. Board-certified teachers will be sought after and will be expected to make a powerful contribution to the quality of education available to America's youth, preparing them to deal with the challenges of the next century.

By retaining in and attracting to the teaching profession high-quality professionals, schools will improve student learning. Professional collaboration among teachers will become a fact of life, as Board-certified teachers assume new roles as instructional leaders and develop supportive relationships with other teachers preparing for Board certification. Schools that effectively use the new professional resources at their disposal should improve the quality of learning for all of their students and should yield graduates who are better prepared to deal with the social, civic, and economic challenges that await them.

These are not small dreams. They will not be easy to achieve. They require the best effort from all of us, and their realization depends on the continued commitment of all of us to the best in American education.

The Paradox of Teacher Assessment

LEE S. SHULMAN
Stanford University

In teacher assessment, we confront a fundamental paradox, a confusion between what is real and what is proxy. We have turned assessment upside-down in thinking about the relationships between tests and the accomplishments they claim to represent. We should strive to set that world upright before we are done.

Any system of quality control calls for a balance between trust and suspicion. We trust the judgments of educational institutions that provide transcripts and degrees to attest to the accomplishments of their graduates. The state and profession set limits on that trust, however, through programs of licensure and certification that protect the public interest. In that spirit, they establish external examinations to evaluate graduates of educational institutions.

Because attestations and examinations are imperfect, most systems of certification attempt to employ both institutional and external sources of evidence. Those educators who know a candidate best may not always be the most dispassionate judges of his or her competence. The most objective external tests and observations may not always yield the most comprehensive or sensitive assessments.

What, then, is the paradox? Tests are at best *proxies* for the direct experience of teaching, interacting with and verifying a candidate's qualifications over longer periods of time and varying conditions. Yet exams are now seen as the ultimate reality, conclusive tests of candidate competence. Individuals who have completed accredited programs of professional teacher preparation are treated in some states as if they have something fundamental to prove. Professional preparation is seen as indirect proof that candidates lack intelligence or proper training — why else would they have "wasted their time in teacher education"? The individual with a liberal arts degree who can pass a relatively brief test or test battery is afforded the same or greater credi-

2U

bility than one who has responsibly pursued a professional preparation program.

Most professional fields — including medicine and architecture — have internship and residency requirements. During the course of these requirements, supervision ensures the candidate's development of important capacities and evaluates the candidate's accomplishment of the desired ends. The licensing and certification examinations in these fields are not perceived as capable of measuring everything of importance. The training programs are expected to teach and to assess in some measure; the external examinations complement the more substantial system of training and induction.

Even in the earliest examinations for teachers, policy leaders recognized that the test scores were not the final word on a candidate's capacities. Note the following instructions given to examiners by the State Superintendent of Instruction in an early Colorado teachers' test (Cornell, 1988; pp. 23-24):

> A high degree of practical success in teaching should be accepted as sufficient reason for issuing a certificate of a higher grade than is warranted by the percentage upon the examination, and inexperience or want of success should lower the grade of the certificate given, while failure as a teacher might be so marked to make it your [the examining county superintendent's] duty to refuse a certificate, whatever the percentage obtained.
>
> I earnestly recommend that certificates of the first grade be given only to teachers who have earned it by success in the schoolroom as well as the examination. I also recommend the addition of ten [points] to the grade earned on Theory and Practice, for the regular reading of some good educational periodicals, or of one or more reliable books on the subject.
>
> Refuse certificates to applicants of whose moral character you have reasonable doubt.

Even 100 years ago, policymakers charged with maintaining teaching standards at the state and county level were admonished to supplement test results with documentation of actual performance in the field. And when a candidate's test scores were inconsistent with his or her experience, greater weight was given to the latter, to what tests could not and cannot measure.

This paper examines the assessment of teachers, the approaches and strategies likely to be useful for evaluating those critically important individuals to whom we entrust the care and instruction of our nation's children. Discussion of this issue, however, should not begin with an inquiry into assessment methods. Method is a second-order question in these deliberations. We must begin instead with a discussion of good teaching, the construct we wish to measure. If the construct is not defined appropriately, there are no measurement methods fancy enough to provide valid assessments.

A Conception of Teaching

Any discussion of teacher assessment requires an adequate theoretical conception of teaching. Teaching always involves both acting and thinking. It typically occurs in connection with specific content and in a particular context. It is interaction with children in classrooms; it is also sitting alone planning at the kitchen table, or talking with parents on the phone. In many systems of staff development and teacher evaluation, the general skills and understandings used to organize and manage classrooms serve as a complete and sufficient definition of expertise. Although these skills are certainly of critical importance, this view of teaching effectiveness is simply inadequate. Any view of teaching effectiveness that fosters a limited view of teaching distorts teacher assessment.

I think of teaching as classroom management and organization — and more. Teaching is not only teacher behavior; it is thought and action with regard to children, purposes, and content in particular contexts. Teacher assessment must measure what, how, and why teachers think about their actions in teaching particular ideas, attitudes, and skills to youngsters in both institutional and community settings.

In turn, this more elaborate concept of teaching demands modes of assessment that go well beyond conventional methods and their underlying assumptions. It requires approaches that present open-ended problems to candidates in a variety of ways. Assessments must confront candidates with problems that require judgments, decisions, choices, and actions that cannot be constrained by the limitations of a multiple-choice test. Those who share a conception of good teaching that emphasizes the importance of varieties of content and context will

not be satisfied by a classroom observation instrument that is so general it can be used to observe a second-grade reading group or an eleventh-grade trigonometry class.

The kinds of understanding and skill that underlie a teacher's expertise distinguish it from the expertise of the subject-matter authority. The teacher deeply understands the content to be learned. The teacher distinguishes between aspects of the content that are absolutely crucial for future understanding and aspects that are more peripheral and less likely to impede future learning if not fully grasped. The teacher recognizes aspects of the content that will be likely to pose the greatest difficulties for pupils.

Teachers also understand that active, constructive, and collaborative student *learning* is the essence of good teaching, not inspired description or energetic demonstration alone. They anticipate persistent preconceptions, misconceptions, or difficulties that are likely to inhibit learning. To form a bridge between students' prior knowledge and key aspects of the concepts to be learned, the teacher invents, borrows, or spontaneously creates powerful representations of the ideas to be learned in the form of examples, analogies, metaphors, or demonstrations. Further, the teacher understands that when students generate such representations themselves, they learn even more successfully. The teacher must therefore be able to distinguish between an inventive contribution from a student that reflects a nonstandard but productive insight into a key idea and another response that communicates confusion and disarray

The teacher understands how to establish a pedagogically meaningful relationship with youngsters. The exemplary teacher creates a relationship around the subject matter to be learned; he or she does not ignore the curriculum and does not establish relationships around social or personal attributes alone. The fine teacher uses his or her understanding to establish relationships with students of diverse cultural backgrounds. Finally, the teacher exhibits and exemplifies a moral or ethical commitment to teach on the basis of the premise that students are capable of learning. Such teachers maintain high expectations for their students without regard to students' prior successes or failures as individuals or as members of a group.

Our conception of teaching attempts to identify the understandings and skills that distinguish an exemplary teacher from an educated pedestrian (defined as a nonteacher who has comparable subject-matter background but no preparation or experience in teaching). It

should also distinguish the teacher of a particular content area from excellent teachers in general. Thus, an outstanding teacher of mathematics would look far more capable when teaching algebra than when teaching *Moby Dick*.

These attributes and accomplishments certainly do not exhaust the types of knowledge and activities that define an excellent teacher. The general pedagogical skills emphasized by Berliner (this volume) are absolutely essential for effective teaching. Yet any system of teacher assessment must focus substantial attention on these content-specific and context-dependent understandings and skills. My colleagues and I began our work with studies aimed at achieving a coherent conception of teaching (Shulman, 1986a, 1987b; Wilson, Shulman, and Richert, 1987). Only after such work was well under way (and much other work by colleagues in the field had been reviewed and analyzed; e.g., Shulman, 1986b) did we ask which approaches to assessment would be adequate to measure teaching competence. Measurement choices must be controlled by pedagogical principles, rather than vice versa. Even though we sensed that we were moving toward kinds of assessment for which little adequate theory or technology existed, the need for teaching to hold the controlling position could not be contradicted. Any system of teacher assessment, however reliable, economic, or efficient, must first and foremost be faithful to teaching.

Toward a Judicious Blend of Assessment Methods

Teaching is such a complex and contextualized phenomenon that any single mode of measurement will fail to assess its practitioners validly. I have proposed that we develop a broad strategy of teacher assessment in the spirit of a marriage of insufficiencies, a complement of imperfections (Shulman, 1987a; 1988). Each method we can imagine for such assessments — the widely used multiple-choice tests, the frequently employed brief classroom visits, my own work on performance assessments and documentation through portfolios — is marred in some fundamental way. The solution does not lie in perfecting the imperfectible, but rather in deploying complementary modes of evaluation that compensate for the most serious deficiencies of their brothers and sisters in measurement.

Indeed, I envision a process that unfolds and extends over time, a strategy of teacher assessment that will combine nearly all of the fol-

lowing elements: written examinations of knowledge and reasoning, both multiple-choice and open-ended; performance assessments in the form of simulation exercises, computer-based problems, and structured interviews; observations of teaching, both direct and via videotape; documentation through reflective portfolios that include samples or exhibitions of student work (captioned and discussed), teachers' plans, student evaluations, recordings of classes, and other artifacts produced in the course of classroom life; and combinations of the above methods. In particular, linking portfolios to subsequent assessment-center evaluations may be useful. What are the strengths of each approach? Where are its insufficiencies?

Written examinations. Tests are used regularly to measure basic skills, knowledge of content, and understanding of professional practice. They permit broad sampling of domains and are relatively economical to use. Because they were originally invented to bring greater objectivity to evaluation, they enjoy high reliability in scoring. Newer forms of essay examinations and experimental methods of employing computers to administer and manage objective tests hold promise for the future. The insufficiencies of written exams lie in their remoteness from the complexities and contexts of practice. Tests currently excel at measuring relatively isolated pieces of knowledge (hence their capacity to sample across wide domains) but fail to tap more integrated processes of judgment, decision making, and problem solving in real-life contexts. Taking a written test is quite different from teaching a class, preparing a lesson, or most other activities in which a teacher is engaged.

Assessment-center exercises. Performance assessment has been used for years in assessment centers for the foreign service, in several medical boards, in the architecture exams, in the California bar exam, and in principals' assessment centers (see Byham, 1986; Aburto and Haertel, 1986). Assessment-center exercises for teaching simulate the real problems and processes of the profession. In my research, we have created 20 exercises in which, for example, teachers are observed as they lay out the plan for a lesson, teach that lesson to a group of new students, and then reflect on the episode and review it critically. Another exercise asks a teacher to analyze a textbook and plan to adapt it for use in the classroom. Our exercises present teachers with a videotape of teaching to observe and critique. They then recommend what they would do under the same circumstances. Another type of

exercise presents examples of student work and asks the teacher to respond to the students' errors and insights (Shulman, Haertel, and Bird, 1988).

In general, performance-assessment exercises can reflect the complexities of teaching more faithfully than a test item does. But they too are insufficient. They are relatively expensive to create, administer, and evaluate. They cannot possibly sample subject-matter content or teaching situations as widely as do conventional written tests. Even though far more realistic than tests, they still cut the teacher off from the actual context of teaching — the school, the students, and the history of teaching and learning they share. How do we design assessments that reflect the actual settings in which teachers do their work?

Classroom observation. Direct observation of teaching is widely employed, especially in the Southeastern states, in evaluation of teachers for permanent licensure. In principle, classroom observations can reflect the full complexity of teaching, but they rarely achieve their potential in practice. The problem of sampling is staggering. Many more classroom visits than have ever been used for evaluation are needed to establish "typical teaching performance" (Stodolsky, 1988). In addition, most methods of direct observation employ the most generic of rating scales, applying the same categories of analysis to primary-grade and to senior-high-school teaching, to a spelling lesson and to a discussion of the quadratic formula. This grows in part out of the emphasis on generic teaching skills that has dominated thinking about teaching in both the research and policy communities. It also justifies the employment of principals or other observers who are not specifically trained in the relevant content areas, grade levels, or contexts to evaluate the teaching they observe.

Observation is an attractive strategy because there seems to be so much potential in watching real teaching in real classrooms directly. But such methods have been disappointing thus far because they fail to tap many of teaching's critical dimensions. Too often, the typical observation method for evaluating teaching has been like photographing the "Mona Lisa" with a black-and-white Polaroid camera, or like tape-recording the most sumptuous performance of *Carmen* with an office dictaphone. There is so much potential in direct observation, but typically so limited a harvest.

Documentation through portfolios. To introduce a connection with the contexts and personal histories that characterize real teaching,

portfolios of various kinds have been employed, with limited success. In licensure and career-ladder programs in Tennessee and Florida, among other states, teachers have been asked to submit lesson plans, attendance records, and other indications of their effectiveness. The resulting portfolios were often too large, too well laminated, and of uncertain connection to the efforts of the individuals whose work they were supposed to reflect. The initial idea seemed a worthy one: collect artifacts that reveal how teachers actually teach. But with insufficient time to design specifications for portfolio contents and inadequate opportunities to create conditions that would promote accurate and relevant documentation procedures, states were unable to carry out the experiment successfully. My research group continues to believe that the underlying notion of documentation is sound and, as I shall discuss next, that the insufficiencies can be overcome or compensated for.

Combining Portfolios and Performance Exercises

We have studied the use of performance assessments for several years and feel that we have made substantial progress in the design and scoring of this new facet of teaching assessment. Yet we remain dissatisfied with some of its limitations, especially with regard to those aspects of teaching in which context and time play a central role. In that regard, we are now investigating the use of carefully structured and well-reviewed portfolios whose specifications are carefully defined. Collaboration in the process at the local level is encouraged. Although there have been problems in the past with portfolios as reliable evidence of teacher competence, they retain almost uniquely the potential to document the unfolding of both teaching and learning over time and combine that documentation with opportunities for teachers to analyze what they and their students have done (see the pioneering contribution by Bird, in press).

We are currently field-testing a program of assessment that combines portfolio development and subsequent assessment of performance. Candidates first spend the better part of a year developing a teaching portfolio. In most cases, each required portfolio entry must include evidence of the teacher's plans and activities (including videotapes of teaching when appropriate), as well as examples of student work. When possible, these required entries extend over time, so that

changes in teaching and learning and evidence of the relationships between them can also be observed.

Through a portfolio, a teacher can record his or her work with an at-risk child over the course of a semester, or document the different kinds of tests and assignments used to assess student progress. A teaching portfolio is ideal for documenting the teaching of writing, complete with the repeated iterations of instruction and feedback so crucial to learning in that area. Individual entries or groups of entries can include (or be followed by) teachers' reflections on their own work and their students' accomplishments.

But a recurrent anxiety has been expressed in regard to portfolios. How do we know that a given set of documents truly represents the work of the candidate? Isn't a portfolio an invitation to fraud, because teachers might borrow the work of others and present it as their own, or purchase "off-the-shelf" portfolios from catalogues? Such concerns have beset the use of portfolios in the recent past.

I begin with a radical premise. A portfolio *should* represent coached performance; portfolio development *should* be an occasion for interaction and mentoring among peers. In our research, we expect every portfolio entry to be cosigned or commented on by a mentor or peer who has participated, helped, reviewed, or critiqued the effort. Much like a doctoral dissertation or a studio project in architecture, the performance represented by the portfolio reflects both the efforts of the candidate and the advice of the instructor. The solution to the nightmare of coaching as a problem is to treat collaboration as a virtue. In medicine, for example, we justifiably assume that the accomplishments of an obstetric resident have been carefully supervised and mentored by more experienced teachers and peers. Documentation of the residency is not discounted for that reason. Why should teachers be treated differently in the documentation of their performances during periods of internship or induction into teaching?

We are currently investigating how to use the assessment center as a follow-up to the portfolio. Anything placed in the portfolio is subject to review in the assessment center, just as a doctoral dissertation is reviewed in an oral examination, or an architect's drawings or scale models are juried in a subsequent design competition. A candidate could be asked to present what has been documented in the portfolio and explain it to examiners. Such presentations by candidates, accompanied by examples of student work, video- or audiotapes of their teaching, and commentaries by mentors or colleagues, could then

become the starting point for about half the assessments that would be conducted. The rest of the assessment-center activity would remain individual exercises, independent of particular experiences unique to any candidate.

Images of documentation through portfolios. In general, the ideal portfolio reflects the accomplishments of teaching. The best portfolio provides evidence of teaching through the documentation of learning — both student learning and teacher learning. It provides not only a description of what the teacher did in particular episodes and over time, but also plentiful examples of student work as that unfolded over time, with as many examples of the teacher's instruction, intervention, coaching, modeling, encouragement, evaluations, corrections, and exhortations as are feasible. In the best sense, a teacher's portfolio should be made up of samples from student portfolios. In addition to examples of student learning that go well beyond the usual end-of-year multiple-choice test results, a good portfolio should include evidence of teacher learning and development over time. Portfolios should become both records and means of support for teacher development.

In considering a special role for documentation through portfolios in the repertoire of approaches to teacher assessment, the following characteristics, attributes, and aspects of the approach are noteworthy:

- It permits the tracking of an extended period of teaching and learning without requiring that an observer be physically present. Stodolsky (1988) points out that stable estimates of teacher behavior require more than a dozen visits to a classroom and sampling across topics and content areas.
- It shifts the focus of teacher evaluation, and thus the unit of pedagogical analysis, from the lesson and lesson fragment to the longer teaching unit, the curriculum, and/or the pupils. Present methods treat an individual lesson as the natural category within which teaching excellence can be assessed. How well are objectives set for *this lesson*? Is an anticipatory set created for *this lesson*? But the most serious embarrassments of contemporary education are not simply the result of poor lesson teaching. The lack of continuity and cumulative learning across days and weeks of lessons is among the most serious failures of our educational system.
- It permits inclusion of data on student outcomes, such as student performances, attitudes, dispositions, and evaluations of teaching,

in closer connection to the actual teaching that preceded and accompanied those outcomes.

- It is, of necessity, content- and context-specific. Documentation is always a record of teaching something in particular to students in some specific setting.
- It can institutionalize norms of collaboration and coaching. Development of approaches to teacher assessment that call for collaboration among teachers to produce the necessary records will lead to greater emphasis on reflection and the evolution of professional ethics regarding good practice.
- It can lead to portfolios that can become the basis for subsequent assessments. These assessments, instead of being based solely on hypothetical examples, can be based on the actual context in which the candidate has worked. Thus, by linking portfolios with performance assessment, we increase the credibility of the former, the validity of the latter, and the utility of both.
- It need not be limited to any particular time period in the teacher's career. The development of a portfolio can begin during the early stages of teacher education and extend through the residency and beyond. Some teachers may take only a year to complete a required set of portfolio entries; others may need (or prefer to take) several years. As in mastery learning, no stigma is attached to the amount of time or number of iterations required.
- Participants in stronger teacher-education programs or more structured induction experiences would have an advantage in portfolio preparation. But it would be a legitimate advantage, tied to the likelihood that they were being better prepared. Similarly, graduates of excellent teacher-preparation programs would emerge with half-full portfolios. They would have an advantage over those who were entering teaching through alternate routes that did not include professional preparation. But this would be a legitimate advantage rather than an artificial, politically legislated one. The ways in which a portfolio is accumulated, reviewed, and reflected on alone and with others could thus serve as a model for teacher education and professional development.
- A portfolio is not part of a teacher's personnel file. A teacher owns and controls his or her portfolio and can add to it and rearrange it regularly, sorting and crafting it into new forms for new purposes,

including personal reflection, professional advancement, district evaluation, and job hunting.

- Documentation may reduce the adverse impact of teacher testing on some minority candidates, because the approach emphasizes actual accomplishments. I base this prediction on my belief that many minority candidates can teach better than they currently test. Documentation may also help to identify another problematic group of teaching candidates to which we devote far too little attention — those who test far better than they teach.

In summary, the virtue of documentation approaches is that the structured portfolio can become a *portable residency*, a set of standards and activities to guide both candidates and institutions during the period of teacher education and induction. Within that framework, teacher preparation and teacher assessment can interact during the critical early years of a teaching career. Assessment using portfolios can thus become more than a new technology for quality assurance; it can become a significant opportunity for the reform of teaching.

The Aims of New Approaches to Teacher Assessment

This paper began with a discussion of paradox. How can we raise teaching standards without emphasizing more stringent forms of external assessment that further erode the credibility of teacher-education programs as sources of attestation? We seek a model of teacher assessment that will ultimately make the institutions of teacher education and teacher induction less dependent on external assessment, less prone to suspicion, and more deserving of trust from the policy community and the society at large.

Sadly, teacher-preparation programs and, in particular, the undergraduate or graduate liberal arts programs with which they are associated, appear to have earned a fair amount of the suspicion they now evoke. Passing grades in content courses cannot be taken as evidence that a future teacher adequately understands the facts and principles of the school subjects in the curriculum. Positive evaluations during student teaching rarely inspire confidence that the novice teacher has demonstrated readiness to teach. One of the implicit messages of this paper is that newer forms of assessment, especially the develop-

ment of documentation through portfolios, can help restore trust in preparation programs by providing visible evidence of what students accomplish.

Ironically, the externalization of assessment could result in less-responsible professional education. Education improves as instruction and assessment are integrated; this was the message of last year's ETS conference, and it is as true for professional education as it is for grades K-12. The emphasis on high-stakes mandatory external assessment for all teacher-education graduates tends to relieve instructional programs of their obligation to evaluate candidates during the course of instruction. Approaches to teacher assessment that would include systematic documentation of the developing capacities and accomplishments of new teachers during their period of preparation could help to convince policymakers that teacher-preparation institutions can be trusted to evaluate their students appropriately.

Assessment can become a powerful tool for reform when it is used to coordinate otherwise disparate, insular efforts for change. Assessment will lead to the enablement and empowerment of teachers when it withers away, when its standards and procedures become integral parts of professional preparation and not external appendages to instruction.

The criteria of validity in such an assessment program extend beyond the usual concerns with reliability and conventional forms of validity. An assessment program must be validated against the degree to which its implementation moves the entire enterprise of which it is a part closer to accomplishment of its social purposes (Cronbach, 1987). We must ask whether a given form of assessment will lead to actions on the parts of teachers and candidates that will make them better teachers, benefit their students and schools, and help move the profession in desired directions.

We cannot afford to be blinded by any particular research program's limited view of the teaching enterprise. Teaching is more than classroom management and organization; more than knowledge of subject matter; more than can be observed solely in a laboratory, or an hour's classroom visit, or an assessment center, or an interview, or surely in a three-hour test battery. Any approach to teacher assessment that lures us into believing that its method alone captures the essence of teaching should be subject to the most skeptical scrutiny.

Members of the educational policy community have been seduced by their own reckless demand for a quick fix in the matter of teacher

standards. And we in the research and measurement communities have meekly acquiesced to their plans, providing them with the instruments they ordered rather than the responsible criticisms they deserve. The time has come to close ranks as a responsible professional community and communicate to our policy colleagues that important and complex problems will not be solved through simple-minded strategies that can be implemented on a large scale overnight.

The National Board of Medical Examiners (NBME) presents a telling example of how modestly the most powerful and influential of professional-standards boards began its work. The first NBME examination was held in Washington, D.C., on October 16-21, 1916. Thirty-two physicians applied for the examination, and only sixteen were deemed eligible, perhaps through the equivalent of some sort of portfolio review. Ten candidates, who had earned their medical degrees between 1907 and 1916, appeared for the examination, which consisted of written, oral, laboratory, and clinical tests. Five passed. (Hubbard and Levit, 1985; p. 163)

What, then, is my dream for an ideal strategy of teacher assessment? Let us arrange a marriage of insufficiencies. Join together both portfolios and direct observation as sources of information about how teachers teach, sources that are specific to the particular context in which a teacher works. And context is important, because the sharing of a history of experiences with a group of children provides the basis for much of any teacher's decisions. Then add a well-crafted set of performance assessments, complete with exercises that follow up on the contents of the portfolios, as well as other exercises that can stand alone. Finally, to ensure broad coverage of the many areas of teacher subject-matter competence and knowledge of student characteristics, supplement these approaches with a new generation of written examinations that are less fragmented and discontinuous than the current crop.

If we can achieve such a program of assessment, in which most of the methods are far more faithful to the practice of teaching than current approaches are, we may be on the threshold of both better teacher assessment and, far more important, better teaching and learning in our schools.

References

Aburto, S., & Haertel, E. (1986). *Study group on alternative assessment methods.* Summary of working seminar. Stanford, CA: Teacher Assessment Project, Stanford University School of Education.

Bird, T. (in press). The schoolteacher's portfolio. In L. Darling-Hammond & J. Millman (Eds.), *Handbook on the evaluation of elementary and secondary school teachers.* Newbury Park, CA: Sage Publications.

Byham, W. C. (1986). Use of the assessment center method to evaluate teacher competencies. Commissioned paper. Stanford, CA: Teacher Assessment Project, Stanford University School of Education.

Cornell, L. S. (1988). *Colorado State superintendent's biennial report.* Denver, CO: State Department of Education.

Cronbach, L. J. (1987). Five perspectives on test validation. In H. Wainer & H. Braun (Eds.), *Test validity.* Hillsdale, NJ: Erlbaum.

Hubbard, J. P., & Levit, E. J. (1985). *The National Board of Medical Examiners: The first seventy years.* Philadelphia, PA: National Board of Medical Examiners.

Shulman, L. S. (1986a). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15(2),* 4-14.

Shulman, L. S. (1986b). Paradigms and research programs for the study of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd Ed.). New York, NY: Macmillan.

Shulman, L. S. (September 1987a). Assessment for teaching: An initiative for the profession. *Phi Delta Kappan,* 38-44.

Shulman, L. S. (1987b). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57 (1),* 1-22.

Shulman, L. S. (November 1988). A union of insufficiencies: Strategies of teacher assessment in a period of educational reform. *Educational Leadership,* 36-46.

Shulman, L. S., Bird, T., & Haertel, E. (April 1988). *Toward alternative assessments of teaching: A report of work in progress.* Stanford, CA: Teacher Assessment Project, Stanford University School of Education.

Stodolsky, S. (1988). *The subject matters.* Chicago, IL: University of Chicago Press.

Wilson, S., Shulman, L. S., & Richert, A. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring Teacher Thinking.* London, UK: Cassell.

A New Generation of Tests for Licensing Beginning Teachers

CAROL ANNE DWYER
Educational Testing Service

Consider the beginning teacher. This is a figure of almost mythical proportions, alternatively heroic and reviled, the object of scrutiny, research, and criticism, defined always by others: sometimes as a problem, sometimes as a commodity, occasionally as the person who is entrusted with our children's minds and psyches, day in and day out, for 180 days of each of their growing years.

But let me tighten the focus just for a moment on this somewhat fuzzy picture. This is a *beginning* teacher. Our beginner has not yet achieved the lofty status of the expert, the master teacher in full command of subject matter and technique. The beginner is in the profession, but not of it. He or she is entrusted with the full responsibility of a teacher from the beginning, but in some important senses our beginner is not yet fully known or fully ready for all of the responsibilities of teaching. In such circumstances one cannot help but feel conflicting responsibilities: to the teaching profession, to the individuals who have worked so diligently to become teachers, and to the children who are to be taught.

To the profession of teaching, our responsibility is to help establish and maintain appropriate standards for teachers. To the aspiring teacher, our responsibility is to recognize his or her investment in time, money, energy, and commitment to becoming a teacher. To the children, our responsibility is to protect them from the ignorant and the incompetent, for the vast majority of children are a truly captive audience. They have no choice but to be in school, and no voice in who will teach them.

How can we contribute to a fair and constructive resolution of these conflicting responsibilities? It is in the context of this basic dilemma that ETS has set out to create a new generation of teacher assessment.

This morning I would like to tell you what we are doing, why we are doing it, how we are going about it, and something of our most daunting and intriguing problems.

The New Assessment and Its Goals

Last year, the trustees and officers of ETS created a unique opportunity for research and development by creating and funding a group of development leaders, who are charged with turning selected broad educational and measurement goals into reality. The aim of my assignment as a development leader is to completely redesign the ETS assessment process used for licensing beginning teachers and, in so doing, to create a new generation of assessment methods. These new methods should take advantage of the latest advances in measurement, psychological and educational research, and technology. This new generation of teacher assessment should also aim to serve the individuals who are assessed as well as the institutions and agencies that make use of information about their performance. The breadth of these goals, of course, presents many technical, policy, and practical problems, but also represents an unusual opportunity to design an assessment with virtually no preexisting conceptual constraints. It is an unparalleled opportunity that carries with it the responsibility to push harder and farther than we can ordinarily envision.

One of the few conditions attached to this assignment is that our basic assessment be available by 1992. I would hope and expect, however, that many of the research and development activities that we undertake would extend beyond this date to take full advantage of opportunities that require a longer time frame. We have now finished our initial data-gathering year, set our preliminary goals, and made our first set of working assumptions, which I would like to share with you today.

A central requirement of this project is that it operate from a sound research and development base and build on both technological advances and the national spirit of educational reform that we are now experiencing. As part of this developmental model, we will be looking at a number of very basic measurement, technological, and educational policy questions. We also expect that the answers to these questions should then have good generalizability to other assessment situations.

Our efforts should also be integrated with, or compatible with, other significant research and development efforts, such as those of the National Board for Professional Teaching Standards and Lee Shulman's Teacher Assessment Project at Stanford University, which concern themselves with the voluntary assessment of advanced teachers and with some very important questions about the nature of pedagogy and the standards to be set for the profession. In addition, the educational policy questions being addressed by the Holmes Group, the experiments by teacher organizations with new roles for teachers, the questions of the induction of teachers into the work force being addressed by the Rand Corporation, and the questions about the essential nature of teaching being addressed by researchers at Michigan State University and elsewhere are all also important sources of knowledge to be integrated into our total effort.

Cooperation among all of the interested groups is important in order not to miss our current window of opportunity. Institutions, like individuals, should be realistic about their own strengths and weaknesses. Our assessments, and teachers themselves, will be better to the extent that we can cooperate in drawing on the best talent in a wide array of disciplines and institutions to design our new assessment. The current and historical strength of ETS is in psychometrics, of course, but we also have a unique strength in taking a solid base of research and theory and "engineering" it for large-scale practical purposes.

Preliminary Shape of the Assessment

We have had to make many working assumptions as we have begun this project. We are assuming that the total assessment will ultimately consist of many parts, as well as other services not directly part of the assessment per se. We must also assume that any innovative assessment we design will meet or exceed current standards for quality and fairness in assessment — innovation does not confer immunity to the difficult questions posed today about these issues.

We are assuming that because this is an assessment to aid in licensing teachers, we should be measuring knowledge and skills that are required for the whole range of teaching situations for which a beginning teacher will be licensed, not just a single teaching job. We also need to consider the academic disciplines with which teachers

will engage their students and the skill required to transform the knowledge of these disciplines so that individual students can benefit from their teachers' knowledge. Our new assessment will thus assess enabling skills, pedagogical knowledge, and subject-matter knowledge, as well as pedagogical-content knowledge of the type currently being explicated by Lee Shulman and his colleagues in the Teacher Assessment Project.

We need to be clear from the very beginning that our goal in this assessment is *not* to predict who will be the best teachers. Such an assessment would need to be tied directly to a specific teaching context and would require reaching agreement on exactly what constitutes excellent teaching. The National Board of Professional Teaching Standards and Professor Shulman and his colleagues are working hard on some of the central problems related to this area. If our activities can be kept compatible as each of our groups continues on its own path, and if we can make clear the underlying continuum of knowledge and skill that links all teachers, then the possibility of mutual benefits in understanding and improving the entire spectrum of teaching exists.

Our primary task for this year is to explore the feasibility of a licensing assessment that is phased to correspond to important milestones in the process of becoming a teacher and utilizes a number of very different assessment methods. We will identify important enabling skills, subject-matter knowledge, pedagogical knowledge, and pedagogical-content skills, and then assess each of these at the appropriate time in the development of a teacher and with the appropriate methods.

Our assessment has three stages that are all part of a comprehensive licensing process. For the first stage of the assessment, we believe that the most appropriate time to evaluate the enabling skills essential to any teacher (such as reading, writing, and mathematics) is very early in the student's undergraduate career, perhaps during the second year of college. We envision a diagnostic assessment that will lend itself well to a computer-based adaptive approach. We also hope to integrate this assessment with technology for individualized career counseling, so that students may learn more about the field of teaching and how it might, or might not, relate to their own interests. We feel that this is best done at a time when those who aspire to teach are still in a position to strengthen their essential skills before graduation, or to make changes in their career plans after consideration of the prospects they

see for success and satisfaction in the career of teaching. This stage of the assessment implies the active participation of the college on the basis of state-determined standards.

The second stage of the proposed assessment, an assessment of subject-matter knowledge and knowledge about teaching and learning, should, we believe, take place immediately on completion of teacher training — however long that may take in the future. Prospective teachers should be able to demonstrate a thorough knowledge of the subjects they wish to teach, and a thorough knowledge of the basic principles of teaching, learning, and human development. We expect that this stage of the assessment will utilize more traditional assessment methods than either of the other stages.

The third stage of our assessment is based on the assumption that an individual's teaching performance — his or her pedagogical skills and pedagogical-content knowledge and skills — needs to be assessed after the individual has had a substantial opportunity to practice teaching. In our vision of assessment for teacher licensing, the development and demonstration of these pedagogical skills depend on actual practice and interaction with students. Therefore, this stage could be linked with the substantial interest on the part of many states in developing beginning-teacher mentoring, intern, or residency programs. It also appears well suited to measurement through performance assessment methods rather than paper-and-pencil testing. It will take place after a teacher is working in an internship or other provisional capacity, but before he or she is fully licensed to teach. We will be exploring a number of possibilities for this kind of assessment and expect that the end result will include a mixture of assessment methods for this stage. We will be looking closely at the use of assessment centers, at systems for in-class observation, and at the feasibility of various types of portfolio documentation of teaching accomplishments.

Some Assessment Alternatives

Our plans for implementing this new assessment call for exploring an extraordinarily wide range of technical and operational alternatives with a broad spectrum of goals, potential benefits, and risks. Some of these alternatives are clearly implied by the three-stage assessment I have just described, such as the extensive use of the capabilities of the personal computer and the use of "live" teaching experiences as the

basis for assessment. I would also like to describe a few of the alternatives I find most exciting, among them some that will probably be the most challenging to achieve.

We will be exploring new ways to present assessment information to the prospective teacher. These will not always seem like "questions" as we know them now, but will be closer to actual teaching or learning situations. Among the models for such new presentation modes is the use of interactive videodiscs to show actual classroom scenes that the viewer can alter via the computer keyboard. Intensive psychometric work will be necessary to reach the technical standards we set for other forms of assessment, but the potential gains in capturing important and complex aspects of the beginning teacher's task make this seem well worth the effort.

We will also be exploring new ways to make what we have learned from the assessment more meaningful to those who are being assessed and to those who must make decisions about prospective teachers. Among these alternatives are providing what can be called diagnostic, or instructionally relevant, information to the test taker, to suggest relative areas of strength or weakness. Such feedback is probably most useful when it is in narrative or verbal form to supplement any numerical information.

We are also grappling with the issue of how to make informed judgments in areas in which there are no single right or wrong answers. Many difficult and important teaching skills involve this kind of problem. Here we will be exploring the use of profiles rather than simple numerical scoring; the feasibility of evaluating these profiles for appropriateness under various conditions rather than classifying them absolutely as "right" or "wrong"; and the possibility of allowing a variety of right ways of responding to particular teaching situations, while identifying certain actions as unequivocally bad practice. Such warning flags might then be used to trigger more intensive assessment for some individuals.

One of the main perceived needs in any new assessment system for beginning teachers is flexibility in administration of the assessment and in its content. I believe great gains in this area are now possible because of advances in technological capability and availability. The need for flexibility is one of the most expensive and difficult to meet, but important for a number of far-ranging reasons. Teaching as a profession is changing with the help of teacher organizations and others. States now differ, and no doubt will continue to differ, in their require-

ments for licensing beginning teachers. This is an unusually active period for research on teaching, and assessments will change as this research proceeds.

By 1992 we will need to have great flexibility about testing modes as well as the content of the assessment itself. Despite the considerable advances made in both the power and the affordability of computers, we still do not have uniformity of access to computers throughout our country. Because of this, some of our assessment will need, at least in the short term, to continue to be offered in paper-and-pencil format rather than solely in a computerized mode. This is especially significant for the stage one and stage two knowledge and skill assessments that I have just described. I also expect that there still will be a continuing, but limited, role for multiple-choice testing in the assessment that I envision. This seems particularly likely in the second stage of the assessment that follows teacher training, when we will be attempting to assess prospective teachers' subject-matter knowledge in certain areas that lend themselves especially well to this format.

Policy and Technical Problems

The assessment that I have outlined for you today presents such a large number of technical problems that it is difficult to choose among them for purposes of discussion, but for the total picture of our new assessment, both our opportunities and our obstacles need to be considered. Here is just a sampling.

The emerging state of research on teaching and the 1992 deadline for implementation of our new assessment are in conflict. A major challenge is how to articulate research and development when research is generating new knowledge at an exponentially increasing rate, and yet that research base is not ready today to be translated into practice. Development must press ahead, trying to remain flexible as research continues to redefine our assumptions. Under these circumstances how can we best define and decide what should be measured?

The assessment of teaching also represents a classic construct validity problem on a truly formidable scale: representing a very complex and ill-defined phenomenon. We will be spending a great deal of time trying to understand what we mean by a teacher who is qualified to be licensed. We will be searching for ways to define this operationally, to approximate it for assessment purposes, and to identify and eliminate

possibilities for bias and other forms of invalidity in our process that would interfere with good measurement and good educational practice. We will take advantage of the concept of utilizing multiple methods for measuring various components of important skills, in hopes of eventually converging on the core problems in assessing the beginning teacher.

Another major problem we face is responding to the need to contextualize the assessment and to bring it closer to specific teaching situations, rather than trying to define a completely generic concept of teaching. Again, this involves some essential questions about the definition of skills necessary for beginning to teach that are not yet resolved. We cannot — and should not for licensing purposes — create separate assessments for every teaching situation, or even every teaching level. But how many can we create, or should we create? What is financially realistic for the developer of the assessment and for the aspiring teachers? We may need in the beginning to design a more complex system in order to help us understand the problems of assessing teachers, then focus on methods of streamlining that process both conceptually and practically.

There are also more policy-oriented problems that we face. Formal assessments are only one part of the long process of developing a corps of effective teachers. Many other evaluative and quasievaluative steps exist in teacher licensing and on-the-job evaluation. Some of these steps are very weak from both scientific and policy points of view, which can increase the likelihood of a number of problems, including some related to group and individual fairness. How can we develop an assessment that has integrity in itself, and yet is well articulated with other steps in the process that have not received the kind of intensive research and development attention that our assessment has?

Another important question for us is to what degree should an assessment process, no matter how soundly based in research and practice, lead the field? There are cogent arguments to be made on either side of this question, but it represents an unusually difficult choice for a project that is committed to innovation as one of its basic assumptions. Can or should this innovation in methods be accompanied by innovation in defining the content of such assessments?

I must also mention the issue of educational equity, and our responsibilities for contributing to it. This project will reflect our deep commitment to address equity issues in assessment and in education more

broadly. Today many minority students still do not receive the quality of education they should. These problems exist in every area of the country and will continue to exist in 1992. These are problems that affect today's students and tomorrow's teachers, and no method of assessment, however innovative, can make them go away. Our new generation of teacher assessment will still reveal educational inequity where it exists, and the implications of this unpleasant message will still be resisted. As an integral part of our new assessment, we must continue our technical and professional efforts to develop an assessment that is itself free from the reality or appearance of bias in any form. Just as importantly, we will need to evaluate, at every step of our research and development process, whether our efforts contribute to helping prospective teachers do the best work they are capable of, both during the assessment itself and in the practice of their profession.

Conclusion

In the licensing of beginning teachers, as in many other situations in which assessment is necessary and controversial, we face difficult decisions that must be made with or without the assistance of formal assessment. I continue to believe that formal assessment can make a very substantial contribution to the fairness and rationality of such decisions.

The task we have set for ourselves in designing a new generation of teacher assessment involves the difficulties of integrating research with practice and values with technology. Such an undertaking necessarily carries with it a high degree of risk, and we face many varieties of potential failure. We can, I am confident, design many innovative and creative assessments, but can we also design assessments that will be widely used and will make a difference to teachers and their students? In 1992, I would be most satisfied with some mixture of success and failure. To my mind the ideal judgment for history to make of this project would be, "They took a lot of risks. They failed with some and learned from their failures, but it paid off for teachers and their students."

Since we are talking today of assessment, I must ask how you will know in 1992 if we have been successful. I hope you will be working on this problem with me for the next four years — what *is* good for teaching, for teachers, and for students, and how will we know it when we see it?

Implications of Studies of Expertise in Pedagogy for Teacher Education and Evaluation

David C. Berliner
Arizona State University

Experts in areas as divergent as chess, bridge, radiology, nursing, air-traffic control, physics, racehorse handicapping, and pedagogy show certain kinds of similarities. Despite their apparent diversity, experts in these fields seem to posses similar sets of skills and attitudes and to use common modes of perceiving and processing information (Chi, Glaser, and Farr, in press; Berliner, 1986). These abilities are not found among novices. Experience allows experts to apply their extensive knowledge to the solution of problems in the domain in which they work. To the novice, the expert appears to have uncanny abilities to notice things, an "instinct" for making the right moves, an ineffable ability to get things done and to perform in an almost effortless manner.

Although we have gained some insight into the differences between experts and novices in various fields, we have only the scantiest knowledge about the ways that one progresses from novice to expert within a field. In part, this is because scientific knowledge about expertise is relatively new. But such research also requires longitudinal studies, and these studies are among the most difficult for which to get support.

Despite the shortage of scientific research in this area, some thoughtful speculation about the ways in which one becomes expert in a particular field is needed, because the planning of instruction for novices and the evaluation of others in a field are inherently related to theories of the development of expertise within the field. If we focus on the field of teaching, then answers to questions about what to teach novices, when to teach it, and how to teach it depend in part on implicit theories about the role of experience in the ability to learn the pedagogical skills, attitudes, and ways of thinking that teacher educa-

tors believe to be desirable. The evaluation of teachers also depends on such implicit theories of development. What one chooses to observe or test for, when one expects to see it, how it should be measured, and the criteria by which successful performance is judged all depend on some notions, perhaps fragmentary, about the development of ability in pedagogy. To make these often implicit and incomplete theories more explicit and complete, I report here on a general theory about the development of expertise. This theory, proposed by Dreyfus and Dreyfus (1986; see also Benner, 1984), provides a way to think about the characteristic stages one moves through as ignorance is overcome and expertise is achieved. My colleagues and I have collected data about pedagogical expertise that is generally supportive of this theory. I will report on some of our findings, acknowledging in advance their tentative nature, and attempt to draw from them policy considerations about teacher education and evaluation.

A Theory of Skill Learning

There are five stages to consider in the journey one takes from novice to expert teacher. We begin with the greenhorn, the raw recruit, the *novice*. Student teachers and many first-year teachers may be considered novices. As experience is gained, the novice becomes an *advanced beginner* (see Bullough, 1989, for a case study of the transition from novice to advanced beginner). Many second- and third-year teachers are likely to be in this developmental stage. With further experience and some motivation to succeed, the advanced beginner becomes a *competent* performer. It is likely that many third- and fourth-year teachers, as well as some more-experienced teachers, are at this level. At about the fifth year, a modest number of teachers may move into the *proficient* stage. Finally, a small number of these will move on to the last stage of development — that of *expert* teacher. Each of these stages of development is characterized by some distinctive features.

Stage I: Novice. This is the stage at which the commonplace must be discerned, the elements of the tasks to be performed must be labeled and learned, and a set of context-free rules must be acquired. In learning to teach, the novice is taught the meaning of terms such as "higher-order questions," "reinforcement," and "learning disabled." Novices are taught context-free rules such as "Give praise for right answers," "Wait at least three seconds after asking a higher-order question,"

"Never criticize a student," and that old standby, "Never smile until Christmas." The novice must be able to identify the context-free elements and rules in order to begin to teach. The behavior of the novice, whether that person is an automobile driver, chess player, or teacher, is very rational, relatively inflexible, and tends to conform to whatever rules and procedures the person was told to follow. Only minimal skill should really be expected. This is a stage for learning the objective facts and features of situations and for gaining experience. And it is the stage at which real-world experience appears to be far more important than verbal information, as generations of drivers, chess players, and student teachers have demonstrated.

Stage 2: Advanced beginner. This is when experience can meld with verbal knowledge. Similarities across contexts are recognized, and episodic knowledge is built up. Strategic knowledge — when to ignore or break rules and when to follow them — is developed. Context begins to guide behavior. For example, advanced beginners may learn that praise doesn't always have the desired effect, such as when a low-ability child interprets it as communicating low expectations. The teacher may also learn that criticism after a bad performance can be quite motivating to a usually good student. Experience is affecting behavior, but the advanced beginner may still have no sense of what is important. Benner (1986, pp. 23-24) makes this point in describing the difference between novice and advanced beginner nurses on the one hand and competent nurses on the other:

> I give instructions to the new graduate, very detailed and explicit instructions: When you come in and first see the baby, you take the baby's vital signs and make the physical examination, and you check the I. V. sites and the ventilator and make sure that it works, and you check the monitors and alarms. When I would say this to them, they would do exactly what I told them to do, no matter what else was going on. . . They couldn't choose one to leave out. They couldn't choose which was the most important. . . . They couldn't do for one baby the things that were most important and then go on to the other baby and do the things that were most important, and leave the things that weren't as important until later on. . . . If I said, you have to do these eight things . . . they did those things, and they didn't care if their other kid was screaming its head off. When they did realize, they would be like a mule between two piles of hay.

The novice and the advanced beginner, though intensely involved in the learning process, may also lack a certain responsibility for their actions. This occurs because they are labeling and describing events, following rules, and recognizing and classifying contexts, but not actively determining through personal action what is happening. The acceptance of personal responsibility for classroom instruction occurs when personal decision making, willfully choosing what to do, takes place. This occurs in the next stage of development.

Stage 3: Competent. There are two distinguishing characteristics of competent performers. First, they make conscious choices about what they are going to do. They set priorities and decide on plans. They have rational goals and choose sensible means for reaching the ends they have in mind. In addition, they can determine what is and what is not important — from their experience they know what to attend to and what to ignore. At this stage, teachers learn not to make timing and targeting errors. They also learn to make curriculum and instruction decisions, such as when to stay with a topic and when to move on, on the basis of a particular teaching context and a particular group of students.

Because they are more personally in control of the events around them, following their own plans, and responding only to the information that they choose to, teachers at this stage tend to feel more responsibility for what happens. They are not detached. Thus they often feel emotional about success and failure in a way that is different and more intense than that of novices or advanced beginners. And they have more vivid memories of their successes and failures as well. But the competent performer is not yet very fast, fluid, or flexible in his or her behavior. These are characteristics of the last two stages in the development of expertise.

Stage 4: Proficient. This is the stage at which intuition and know-how become prominent. Nothing mysterious is meant by these terms. Consider the microadjustments made in learning to ride a bicycle — at some point, individuals no longer think about these things. They develop an "intuitive" sense of the situation. Furthermore, out of the wealth of experience that the proficient individual has accumulated comes a holistic recognition of similarities. At this stage, a teacher may notice without conscious effort that today's mathematics lesson is bogging down for the same reason that last week's spelling lesson bombed. At some higher level of categorization, the similarities between disparate events are understood. This holistic recognition of

similarities allows the proficient individual to predict events more precisely, since he or she sees more things as alike and therefore as having been experienced before. Chess masters, bridge masters, expert air-traffic controllers, and expert radiologists rely on this ability. The proficient performer, however, while intuitive in pattern recognition and in ways of knowing, is still analytic and deliberative in deciding what to do. The proficient stage is the stage of most tournament chess and bridge players. But the grand masters are those few who move to a higher stage, to the expert level.

Stage 5: Expert. If the novice, advanced beginner, and competent performer are rational and the proficient performer is intuitive, we might categorize the experts as often arational. They have both an intuitive grasp of the situation and a nonanalytic and nondeliberative sense of the appropriate response to be made. They show fluid performance, as we all do when we no longer have to choose our words when speaking or think about where to place our feet when walking. We simply talk and walk in an apparently effortless manner. The expert safety in football, the expert martial artist in combat, the expert chess master, and the expert teacher in classroom recitations all seem to know where to be or what to do at the right time. They engage in their performance in a qualitatively different way than does the novice or the competent performer, like the race-car driver who talks of becoming one with her machine or the science teacher who reports that the lesson just moved along so beautifully today that he never really had to teach. The experts are not consciously choosing what to attend to and what to do. They are acting effortlessly, fluidly, and in a sense this is arational, because it is not easily described as deductive or analytic behavior. Though beyond the usual meaning of rational, since neither calculation nor deliberative thought is involved, the behavior of the expert is certainly not irrational. The writings of Schon (1983) about knowledge in action characterize the behavior of the expert practitioner.

Experts do things that usually work, and thus, when things are proceeding without a hitch, experts are not solving problems or making decisions in the usual sense of those terms. They "go with the flow," as it is sometimes described. When anomalies occur, things do not work out as planned, or something atypical happens, they bring deliberate analytic processes to bear on the situation. But when things are going smoothly, experts rarely appear to be reflective about their performance.

I bring this theory before you because I believe it has heuristic value for thinking about educating and evaluating teachers. Furthermore, it is reasonably well supported by data my colleagues and I have collected over the last few years, although we did not know about the theory when we started our work. In our studies of pedagogical expertise (and other studies that may be cited), the subjects were categorized and labeled differently than those used to describe the five stages of development in the general theory presented above. In an attempt to reduce confusion and synthesize findings across studies, I will translate the descriptive terms used in the original research into the language used to describe the five stages of expertise, and I will state the findings we think have emerged in the form of principles. This should aid us in interpreting research in this area but cannot serve as a substitute for careful review of the procedures used and findings reported in the original documents.

Findings and Implications

1. *There are differences in the ways that teachers at various levels of experience and expertise interpret classroom phenomena.* Because of a lack of experience, those near the novice end of the developmental continuum can be expected to have trouble interpreting events. Until episodic knowledge is built up and similarities can be recognized across contexts, confusion may characterize the interpretations of classroom phenomena made by novices and advanced beginners. Experts are more likely than those with less ability to discern what is important from what is not when interpreting classroom phenomena. And we should also expect that experts will show more effortless performance and rely more on experience for interpreting information. We obtained data supportive of these ideas in some of our studies.

   A. *Variation in what they direct their attention to and confusion in interpreting classroom information result in the inability of novices and advanced beginners to make sense of what they see and hear.* In one experimental situation, experts, advanced beginners, and very raw novices watched three video screens, showing the same lesson from different camera angles (Sabers, Cushing, and Berliner, under review). The three monitors played simultaneously, creat-

ing a complex auditory and visual environment. The advanced beginners apparently experienced difficulty in making sense of their classroom observations and in providing plausible explanations about what was occurring within the classroom. They often made contradictory statements about what they had observed. For example, when asked to describe the learning environment in the classroom, they answered as follows:

Advanced Beginner: It looked . . . I wouldn't call it terribly motivating. It was, well, not bored, but not enthusiastic.

Advanced Beginner: Very positive as well as relaxed. Very positive . . . it's good to be able to focus [student] energy into a group situation, yet at the same time, accomplishing the work that they need to do for the class and also lending to the relaxed feeling of the classroom.

And when the advanced beginners were asked to comment on the students' attitudes toward this class, we heard these comments:

Advanced Beginner: It didn't look like it was a favorite class for most of them. One boy looked kind of like, "Oh no, it's not this class again." They didn't look overwhelmingly enthusiastic to be there.

Advanced Beginner: They seemed pretty excited about the class, excited to learn and a lot of times it's hard to get students excited about science, but this teacher seems to have them so that they are excited about it. They're willing to work and they want to learn.

Such contradictory interpretations were common. Because the advanced beginners experienced difficulty in monitoring all three video screens at once, they seemed unable to make much sense of what they saw.

The novices in this study were much less familiar with classroom events than were the advanced beginners. Perhaps that is why they appeared even more overwhelmed. Many of the novices expressed difficulty or an inability to monitor all three video screens at once. Generally, they appeared able to focus on and make sense of only one video screen. Since this limited their observations, they also made errors and contradictory statements when they were asked about specific events.

Experts, on the other hand, did not demonstrate any confusion or difficulty in making sense of their classroom observations when presented with the videotapes. The experts responded effortlessly and fluidly. They not only made more comments about what was happening, but their comments were more detailed and descriptive than those of the other two groups. The experts appeared comfortable both describing what they observed and interpreting those events on the basis of their experience in classroom instruction and management. Part of their proficiency was due to the fact that experts were able to monitor both the teacher and students more accurately than advanced beginners or novices were. Given their ability to monitor three video screens simultaneously for both visual and auditory cues, experts seemed less confused than either the advanced beginners or the novices and thus were better able to interpret, evaluate, and make hypotheses about classroom events. In contrast, some advanced beginners and almost all the novices gave step-by-step descriptive accounts of what was happening, as though they were announcing what they were viewing to someone who could not see the screen. In this respect they were reminiscent of radio announcers reporting an athletic event. The verbal reports of advanced beginners and novices indicated that they had no difficulty perceiving, but considerable difficulty in comprehending, the events and behaviors that they had listened to and viewed. Lacking in their verbal reports were the inferences, predications, conclusions, evaluations, and suggestions that were reported frequently by the experts, characteristics discussed in more detail below.

Even in simpler environments, the novices and advanced beginners had trouble making sense of what they saw. In another of our studies (Carter, Cushing, Sabers, Stein, and Berliner, 1988), our subjects went through a series of slides depicting a class period of high-school science or mathematics instruction. The subjects held a remote control and were told to go through the 50 or so slides at their own pace, stopping to comment on any they found interesting. The advanced beginners and novices showed the same kinds of contradictions in their interpretations of what they saw as we found in the videotape study. That is, one advanced beginner or novice might say, "Everything looks fine; they're all paying attention," and another subject from these

groups might say, "It looks like they're starting to go off task; they're starting to drift."

Not only was the information reported by the novices and the advanced beginners in the slide study often contradictory, but there was also no particular pattern in what they stopped to comment on. In contrast, a pattern was noted among the experts that was quite different. The experts, more often than other subjects, found the *same* slides worth commenting on and had the *same* kinds of comments to make. For example:

Expert: [Slide 5] It's a good shot of both people being involved and something happening.

Expert: [Slide 5] Everybody seems to be interested in what they're doing at their lab stations.

Expert: [Slide 5] Everybody working. A positive environment.

Expert: [Slide 51] More students with their books closed, their purses on their desks, hands folded, ready to go.

Expert: [Slide 51] Must be the end of class and everybody is getting ready for the bell to ring.

This greater uniformity in the interpretations of the experts is particularly noteworthy. It means they have learned to pay attention to some of the same things and to interpret visual stimuli in the same way. This similarity in *what* is attended to and *how* it is interpreted is what we hope for when we visit an expert ophthalmologist or an expert tennis player for advice. Novices and advanced beginners — people in the early stages of skill acquisition — simply have not acquired enough experience for that. Their behavior, in contrast to that of the experts, is characterized by more variation in what is attended to and by more confusion about how information is to be interpreted. William James' characterization of the neonate's world as one of "booming, buzzing confusion" appears to be characteristic of the neophyte's world as well.

B. *There are differences in the ways that novices, advanced beginners, and experts perceive classroom information, with experts showing an ability to distinguish the typical from the atypical occurrence and responding primarily to the latter.* The general theory presented

above suggests that experts will not pay much attention to things that, in their judgments, are going smoothly. In fact, the expert appears rather unreflective as long as no problems are perceived. This is another reason that the expert appears so effortless at work. Experts apparently pay attention to fewer things than do others whose skills are less developed. This does not mean that experts see or hear less, only that they choose to process less of what they encounter. In fact, there is reason to believe that they actually see and hear more than do novices (Sabers, Cushing, and Berliner, under review; Peterson and Comeaux, 1987). In two of our studies, we saw some evidence of this tendency for experts to focus on the atypical.

In one study we asked our subjects to examine a great deal of information about a class whose regular teacher resigned in the fifth week of a semester. The simulation was designed to inquire into how these expert, advanced beginner, and novice teachers prepared for taking over the class. Experts acted differently from the others. While novices and advanced beginners worked diligently to learn all they could about the various students in the classroom, their achievements, problems, home life, and so forth, the experts invested very little mental effort in this activity. They seemed to merge information about students into a "group picture" that they defined as more or less "typical," "normal," or "usual." For example:

Expert: Especially when I start fresh, I start from a clean slate. . . . I like getting a little background on the students — that there are going to be severe problems or someone may need special attention on certain things, you know, learning areas, but, in general, it's a conglomeration of the students. I like to learn from them and develop my own opinions.

Expert: It was a typical classroom, some problem kids that need to be dealt with. And you have to take that into consideration when you're developing some kind of plan for them. There were the bright kids that were highly self-motivated. There were your shy kids. It was a typical class.

Expert: I didn't read the cards. I never do unless there's a comment about a physical impairment such as hearing or sight or something I get from the nurse. I never want to place a judgment

on the students before they start. I find I have a higher success rate if I don't.

The belief that it is atypicalness that is attended to was supported in a different context, in the study in which slides were used as stimulus materials. Experts tended to describe one of the slides differently than did novices or advanced beginners. For example:

Expert: Not a typical row, not a typical group for conversation. . . . Part of them were facing this way and part of them were facing the other direction so it wasn't like a typical situation.

Experts, apparently, have images of how things ought to be. Students or classrooms that appear to be the way they are supposed to are less likely to be attended to — perhaps more likely to be ignored. Experience seems to change people so that they literally "see" differently, either by noting atypicalness quicker (the experts seem to have quicker pattern-recognition ability) or by simply not seeing certain ordinary things. Surely that is functional. In any domain of expertise one must learn through experience, perhaps thousands of hours of experience, what is worth attending to, particularly because of the severe biological limits humans have for processing information.

C. *There are differences in the ways that novices, advanced beginners, and experts perceive classroom information, with experts showing greater ability to pick out what is important from what is not.* The theory discussed above suggests that this is one of the characteristics that distinguishes a neophyte from those at higher levels of development. We have evidence about that from our study of visual processing of classroom information. We showed a slide of a classroom scene very rapidly and asked our subjects what they saw. The responses of the novices and advanced beginners to the brief exposure to a slide were clearly descriptive and usually quite accurate.

Novice: A blond-haired boy at the table, looking at papers. Girl to his left reaching in front of him for something.

Advanced Beginner: [It's] a classroom. Student with back to camera working at a table.

Advanced Beginner: A room full of students sitting at tables.

In contrast to these literal descriptions typical of novices and advanced beginners, some of our expert teachers did not give mere descriptions; their perception was directed more by their experience.

Expert: It's a hands-on activity of some type. Group work with a male and female of maybe late junior-high-school age.

Expert: It's a group of students maybe doing small-group discussion on a project as the seats are not in rows.

For experts, the information that was often deemed important was information that had instructional significance, such as the age of the students or the teaching/learning activity in which they were engaged. Blond hair and posture were simply not important. In this sense our experts were acting like experts in other fields do, responding not to the literal characteristics of a situation, but only to those aspects of a situation that are important for their work. The experts have apparently learned to discriminate between what is and what is not important.

In a replicate of this study that generally supported the findings from our research, Nelson (1988) concluded that expert physical-education teachers "interpret more cues from limited information, enabling them to impose more meaning and make more inferences from these cues." On the basis of both studies, we might think of characterizing the processing of classroom information by novices and advanced beginners as "bottom-up," since it is driven by the surface properties of the stimulus materials. In contrast, experts show "top-down" processing, no longer perceiving literally, using instead their experience to interpret the situation.

D. *There are differences in the ways that novices, advanced beginners, and experts evaluate teaching performances.* Unlike the novices and the advanced beginners, who either showed restraint or lacked ability in making evaluative comments, the experts seemed to combine interpretation with evaluation of the events and behaviors they viewed. For example, when viewing a videotaped lesson, experts made the following comments:

Expert: In the monitor in the middle, it might have been a good idea to start out class with measuring the height of these plants that they're growing while roll is being taken, so that you're not wasting or having a bunch of dead time at the end of the class.

Expert: Again, viewing the middle monitor, I think there is an indication here of the type of structure of this classroom. It's pretty loose. The kids come in and go out without checking with the teacher.

In one of our simulation studies, subjects received a packet of information about a class they were to take over, including student-information cards, tests taken by the students, homework turned in, and so forth. Experts again appeared more willing to make evaluative comments then did novices and advanced beginners. For example, about the multiple-choice tests used by the previous teacher, Expert 4 said, "For a learning experience those things are worthless." And about the homework that had been assigned by the previous teacher the same expert said, "That type of homework . . . isn't working. . . . That's just busywork wasting everybody's time." The expert physical-education teachers used by Nelson (1988), in a replicate of our slide-interpretation study, and the experienced (at the competent level and above) teachers of social studies used by Peterson and Comeaux (1987) provided the same kinds of evaluative comments as the experts in our studies. In all three of these investigations, it was independently noted that the interpretations and evaluations of classroom phenomena by the experienced or expert teachers were informed by their extensive experience, making their verbal responses appear richer and their thinking more principled.

An analogous situation to this increased frequency of evaluative comments and reliance on personal experience by expert teachers is found in studies of expert readers. In that research tradition, the ability to evaluate as one decodes or interprets prose and relate what is read to one's existing knowledge base are considered sophisticated and useful metacognitive skills leading to better comprehension. In an analogous manner, expert teachers trying to comprehend information about classrooms appear to have the same cognitive (or metacognitive) skills.

E. *There are differences in the assumptions, predictions, and hypotheses made by novices, advanced beginners, and experts as they interpret classroom phenomena.* The general theory of skill development presented above suggests that experience gives experts more capability for making better assumptions and hypotheses about classroom phenomena and student behavior, assumptions and hypotheses that are not obvious to novices. Because experience leads to recognition of similarities, what is learned is the probability that certain events or stimuli are associated with certain other events or stimuli. In our studies, experts did seem to predict or hypothesize more than did novices or advanced beginners. For example, in the videotape study, experts naturally and confidently made predictions:

Expert 5: I'm looking at the left monitor. . . . I think that this is a part of a continuing activity from the day before probably, because they know exactly what they're doing without any instructions from the teacher.

Expert 7: On the left monitor, the students' note-taking indicates that they have seen sheets like this and have had presentations like this before; it's fairly efficient at this point because they're used to the format they are using.

Expert 2: Left monitor again. . . . I haven't heard a bell, but the students are already at their desks and seem to be doing purposeful activity, and this is about the time that I decided they must be an accelerated group because they came into the room and started something rather than just sitting down or socializing.

In fact, just as this expert inferred, we had recorded an accelerated group of students. It was a science classroom for students identified as GATE (Gifted and Talented Education) students. This tendency of experts to predict, hypothesize, or assume while attempting to interpret was also found in our study of the interpretation of classroom information presented on slides. When talking about what they had seen, experts reported:

Expert: Students were not seated in the traditional type of seating arrangement — one that would normally be used for a lecture-type style of teaching. So from the seating arrangement I

assumed that they must have been involved in some activity other than a traditional type of a lecture.

Expert: There aren't a whole lot of humorous math problems so I *assumed* a couple of the students must have been talking — from their facial expressions — about something other than the assignment.

Expert: I *assumed* it was the teacher's desk because [it] was faced a different way from where the students' desks were faced and because of where it was placed in relationship to the chalkboard.

Assuming, hypothesis-making, and predicting were much more prevalent in the language of experts than in that of novices or advanced beginners. Nelson (1988) also reported that expert physical-education teachers, rather then novices, drew more inferences from the cues available in a series of classroom slides. In addition, we confirmed this characteristic of expertise in yet another study, one that focused on students, not classrooms. In this study, experts, novices, and advanced beginners talked out loud as they thought about how students would answer particular items from the National Assessment of Educational Progress (Stein, Clarridge, and Berliner, in preparation). The experts, novices, and advanced beginners differed in their predictions about the student cognitions used in answering an item. Experts seemed to have a fund of knowledge about the way students thought and how those thoughts interacted with the content of the specific mathematics or science items. In addition, the experts seemed able to think through the misalgorithms that students might apply to solve a particular problem. The experts had more experience dealing with student errors and therefore were able to predict what types of errors students might make. Novices and advanced beginners rarely discussed the issue of misalgorithms. The inability to predict the kinds of errors that students will make in particular situations is, no doubt, a major deficiency of the beginning teacher. If the tough-to-teach topics cannot be distinguished from the easy-to-teach topics, and no taxonomy of error types has been built up through experience, a teacher is likely to teach in an inappropriate way. The ability to predict how students will think and err is fundamental to a diagnostic-prescriptive form of teaching. Such knowledge is generally unavailable to novices.

In another research tradition, the study of reading, scholars have arrived at a similar conclusion. Expert readers, compared to those who have difficulty in comprehending reading, have developed metacognitive skills that allow them to predict events and outcomes in the stories they read. Such predictions increase comprehension. Our studies of pedagogical expertise confirm these beliefs.

F. *There are differences in the frameworks used by experts, advanced beginners, and novices as they interpret classroom information.* We have already mentioned that experts seem to rely more on personally experienced episodic knowledge to interpret information about classrooms. They seem to bring to a situation more comprehensive schemata, well grounded in personal experience, enabling them to make better sense out of what they see and hear. Peterson and Comeaux (1987), Nelson (1988), and Leinhardt and Greeno (1986), among others, have noted this characteristic, and investigators of expertise in other areas have commented on it. (See Chi, Glaser, and Farr, in press.) In addition, a framework unique to experts in pedagogy was suggested in our study of interpretation of classroom activities from slides. In this study, experts demonstrated how they organized their viewing around the concept of work. They commented on "students *working* at the blackboard," "students *working* independently," "teacher looking over a person working in lab," etc. Work appeared to be a salient organizing concept for the experts. This should not be a great surprise, since the experts gain their reputations, in part, because their students win science fairs, achieve well on state tests of mathematics, get high scores on the SAT, and so forth. That is, students in their classes work!

For the most part, the novices' descriptions of the classrooms did not mention student work. In fact, students and their work involvement appeared to be no more salient to novices than were the physical surroundings of the classroom, the equipment, the windows and desks, and the charts on the wall. Apparently, *all* of the visual stimuli presented had equal information value. Novices, lacking organizing frameworks or complete classroom schemata for interpreting classroom information, appeared unable to distinguish what was important from what was not, a characteristic that was commented on above. The inability to

interpret information due to incomplete schemata and a lack of organizing concepts results in an inability to distinguish between the forest and the trees, a characteristic others have noticed when studying those in the early stages of skill development (Lesgold, 1984).

Support for our hypothesis that experts use an organizing framework with "work" as a central concept comes from other sources. Nelson (1988), in her comparison of expert and novice physical-education teachers, also reports that a focus on student behavior differentiates experts' from novices' interpretation of classroom events. Housner and Griffey (1986), using novice and experienced (i.e., competent, proficient, and perhaps expert) physical-education teachers, found that the experienced teachers focused more on student performance than the novices did. Stein and Berliner (1988) reported on experts' thoughts during interactive instruction, comparing them to novices, advanced beginners, and competent or proficient teachers. Once again the experts focused more frequently on student academic work and cognitions then did the other groups. Finally, in another study from our program of research, Clarridge (1988) found that the teaching performance of experts, in comparison to advanced beginners and novices, focused more on student classroom activity; i.e., classroom *work*. It appears that the framework used by experts for interpreting classrooms is one in which student productivity plays a central role.

In summary, from a developmental perspective, there is reason to believe that when interpreting classroom phenomena, those in the early stages of the development of skill in pedagogy:

- Demonstrate more variability in what they pay attention to and in how they interpret what they see and hear.
- Perceive differently from those in the later stages, showing an inability to separate ordinary and typical occurrences from those that are not.
- Perceive differently from those in the later stages, showing an inability to separate what is important information from what is not.
- Do not evaluate the information they receive about classrooms as frequently as do those in the later stages of development, nor do

their commentaries provide evidence of principled thought as frequently or rely on rich personal histories as often.

- Do not predict, hypothesize, or make assumptions as frequently as do those in the later stages of development.

- Do not use the same frameworks for interpreting classroom information as do those in later stages.

Interpretive ability is not the only area in which we uncovered differences between expert, advanced beginner, and novice teachers. Two other areas in which differences appeared are also worth noting here.

2. *There are differences in the use of classroom routines by teachers at various levels of expertise and experience.* The effortless and fluid performance that often characterizes the experts' performance may be due, in part, to their use of routines. Adherence to routines by teachers and students makes classrooms appear to function smoothly. In studying elementary-school mathematics lessons, Leinhardt and Greeno (1986) compared an expert's opening homework review with that of a novice. The expert teacher was found to be quite brief, taking about one-third less time than the novice did. This expert was able to pick up information about attendance, about who did or did not do the homework, and was also able to identify who was going to need help later in the lesson. She elicited correct answers most of the time throughout the activity and also managed to get all the homework corrected. Moreover, she did so at a brisk pace and never lost control of the lesson. She also had developed routines to record attendance and to handle choral responding during the homework checks and hand-raising to get attention. This expert also used clear signals to start and finish the lesson segments. In contrast, when the novice was enacting an opening homework review as part of a mathematics lesson, she was not able to get a fix on who did and did not do the homework, she had problems taking attendance, and she asked ambiguous questions that led her to misunderstand the difficulty of the homework. At one time the novice lost control of the pace. She never did learn which students were going to have more difficulty later in the lesson. It is important to note that the novice showed lack of familiarity with well-practiced routines. She seemed not to act in habitual ways.

Students, therefore, were unsure of their roles in the class. Brooks and Hawke (1985) provide similar data about the differences in routinization of procedures in their comparison of an expert and novice teacher.

In one of our studies (Carter, Sabers, Cushing, Pinnegar, and Berliner, 1987), experts, advanced beginners, and novices talked about what they would do first when taking over a class that had been running for five weeks. The experts, more so than the others, made mention of the need for routines:

Expert: I have to be organized before I can be comfortable. It can't be chaotic. I can't run a class when I don't know what's going on. I've got to be with it, what's going on, and have a routine set up for the kids to respond to, when they know what's expected. . . . It's like training a two-year-old. They have to know what's expected. So I put a lot of emphasis at the start of the year on training the students.

Expert: Okay, the first day I had in mind that I'm going into a brand-new class. They possibly don't know me at all. I would quickly go over what I would expect for rules and expectations, raising their hands before talking, courtesy to teachers and students. . . . I'd give them some rules to follow to get us started.

In this situation the experts, more than the novices, recognized the need to impose their own routines on the classroom, to automate procedures so there would be order. This allowed them to teach in the manner to which they were accustomed.

A. *When experts do not have a chance to develop their routines for working with students in the classroom, they may appear to lose some of their expertise.* In another of our studies (Berliner, 1988), we had experts, novices, and advanced beginners teach a short lesson on probability to about 15 high-school students. The experts were very unhappy about their participation in this task, in part because the students were not the ones they taught regularly. Their own students were trained in routines to make the classroom run smoothly. One expert, reflecting on what was wrong with the task, said:

Expert: My expectations when a kid comes into my classroom for math is that he has pencil and paper ready at all times, because I

make them take notes, just as you do in social studies. They have practice problems. And this is kind of tough 'cause I don't know what was the routine these kids were used to, you know?. . . You know, with the kids that are used to your routine, you can stand up and talk for 15 or 20 minutes, and by your questioning techniques and by having them work with guided practice at their desks [you keep them working]. But these kids didn't know me, and they didn't know the way that I operate, that all are supposed to participate, and why, and that they're all supposed to be on task [constantly].

The well-practiced routines of expert surgeons, ice skaters, tennis players, and concert pianists (Bloom, 1985, 1986), no less than master teachers, are what give the appearance of fluidity and effortlessness to the performance of experts. Bryan and Harter (1899) noted this a long time ago in their classic studies of the development of expertise in telegraphy. They said a person must acquire a system of habits that routinizes, systematizes, or makes automatic many of the person's responses in his or her chosen line of work. A person who has done that "is master of the situation. . . . Automatization is not genius, but it is the hands and feet of genius." What looks to be so easy for the expert and so clumsy for the novice is the result of thousands of hours of experience, during which routinization of many procedures has been acquired. An expert who is put into situations in which well-honed routines cannot be used may appear clumsy. Routinization is certainly not expertise, but as noted 90 years ago, it may well be the hands and feet of expertise.

3. *There are differences in the emotionality displayed by teachers at various levels of expertise and experience.* When the developmental stage of competence is reached, it is said to be accompanied by a qualitatively different kind of emotionality and sense of responsibility for the work of the performer. We have some evidence for that, obtained in a curious way, in the study in which experts, advanced beginners, and novices planned and then taught a lesson (Berliner, 1988). The novices in that study were quite happy about their performance, although we did not rate it highly. Advanced beginners were generally affectless in describing their experience. They had a task to do and they did it. The experts, however, were quite angry about their participation in the task and disappointed about their performance.

In retrospect, and on the basis of our interviews, it appears that we had inadvertently taken away some of the experts' edge. First, we had created an artificial teaching situation. Second, according to their standards, they did not have enough time to prepare the lesson. Third, the students were not trained in the routines that make the experts' classrooms hum. One expert expressed his anger by walking out of the study. Another stopped in the middle of the lesson and had to be coaxed to continue. One started crying during the playback of her videotape. All were upset. Two weeks after the study, one expert, when asked what she remembered of her experience, said:

Expert: I just remember it as the worst experience in my entire life, and I was depressed. . . . The things that stick out in my mind are the negative things. I remember just being frustrated the whole time I taught the lesson. . . . I don't like what happened. I've been real depressed and down [since then].

Other comments by experts were about their feelings of uncomfortableness, stress, terror, and so forth. In this situation, advanced beginners and novices were virtually untouched at any deep emotional level, but our experts were affected deeply. In addition, they felt that in some way they had let us down — their sense of responsibility played a part in their feelings. Expert teachers, apparently like other experts, show more emotionality about the successes and failures of their work.

Summary

A growing body of literature is documenting the ways in which individuals at different levels of experience in classroom teaching and other fields differ in their interpretive abilities, their use of routines, and the emotional investment that they make in their work. From this one can extract a general principle, namely, that very important qualitative differences exist in the thinking and the performance of novices, experts, and all those who fall between these two points on the continuum. The developmental sequence involved in the acquisition of expertise, however, is not yet as clearly described. The five-stage theory of the development of expertise presented above is intended to help us think more about that issue and is well supported by data that

were collected for other purposes. Since a developmental theory is necessary to plan instruction and assessment, some interesting implications may be derived from the use of this model and the supporting set of data.

Implications for Teacher Education and the Assessment of Teachers

Short statements about the implications of this research on pedagogical expertise follow. These are presented as informed speculations to help those who teach and test teachers explore the complex issues involved.

- The degree of emotionality displayed by expert teachers who were removed from their classes and asked to work in a simulated classroom environment was caused, in part, by our removal of the sources of their expertise — adequate planning time, knowledge of the students, personal history with the class, reputation in the school, routinized performance, etc. An implication is that tests of expert teachers may identify many false negatives — those who are actually quite expert in their classrooms, but do not perform expertly when tested in simulated situations outside of their classrooms. Expertise, it would seem, is really quite fragile because it is highly contextualized and may not transfer well. The certification of competence, proficiency, or expertise in pedagogy may have to rely more on classroom observations and reputational measures than on simulations and paper-and-pencil tests.
- Advocates of initial licensure tests for teachers may have to consider that all they can realistically tap is pedagogical knowledge derived from preservice programs of teacher education, and not pedagogical knowledge derived from field experience. Novices need extensive time in the field to make sense of classroom life. The typical student-teaching program does not provide that experience. Perhaps licensure tests should be administered in two parts. The first part would be given at the completion of preservice education and would include items on history, philosophy, subject-matter knowledge, research on teaching, professional issues, and other aspects of the preservice curriculum. The second part might be given at the end of two years of employment and would focus on management,

organization, and similar pedagogical issues. Since answers to such questions are highly dependent on context, this test would be scored by peers who bring to the evaluation notions of reasonable practice. Sophisticated responses to questions about classroom management and organization and the transformation of curriculum into pedagogically useful forms are, perhaps, more than novice teachers can deliver.

- In many of the studies cited, the novices were individuals who enter teaching by alternative certification. Some were individuals with content knowledge, an interest in children, no education courses or formal teaching experience, and a desire to be licensed. Such novices were clearly ignorant about classroom processes. Their interpretive skills were poor. Content knowledge did not substitute for pedagogical knowledge, which advanced beginners acquired through their student-teaching experience as well as during their first year of teaching.

- The theory and data associated with the development of expertise lead us to believe that the real goal of the first-year teacher, entering through traditional or alternative routes, is that of muddling through until it all starts making sense, and until some of what is required to run the classroom can be routinized. From this perspective it seems odd to expect the novice teacher to have responsibility for a full teaching load, with responsibility for managing and teaching the same number of children as the competent, proficient, or expert teacher. Even more worrisome is that the competent, proficient, and expert teachers sometimes take the classes and students that are easiest to teach, leaving for the novice and advanced beginners the most difficult. This is a sure way to keep the dropout rate for teachers in their first five years of teaching as high as it is currently.

- The needs of the first-year teacher have been made public in recent years, and the response has often been to provide a mentor for that teacher. The intent was to choose mentors from the ranks of the expert teachers. Two issues arise from the research cited above. First, the expert pool, despite the fancies of some, is likely to be quite small. Furthermore, the competent teacher may be able to articulate as well as or better than the expert the issues of pedagogy that must be addressed by the novice and the advanced beginner. The proficient performer has been described above as intuitive, and the

expert as arational. Neither of these are inherently desirable qualities for a mentor-teacher. Thus, we conclude that mentors need not be experts themselves; rather, they must be articulate analysts of teaching.

Although the expert teacher may not be the ideal mentor, experts can be very good models. A classic study of expertise in chicken sexing (Lunn, 1948) makes this point abundantly clear. It is virtually impossible to explain to another person how to sex a chicken. But if you stand next to an expert chicken-sexer for three months, imitating the moves and remembering the decisions, you will be able to sex 1,000 chickens per hour with 98 percent accuracy. Being a good model and being a good mentor are different, and we may want to differentiate these roles more clearly in teacher education.

- It is customary in one of the largest districts in my state to require a full training program in elements of effective instruction for all first-year teachers. Given the theory of skill development described above, there is good reason to believe that in-service education programs dealing with sophisticated issues of classroom instruction and management may be less effective than desired if provided during the first year of teaching. That is a time when receiving emotional support, learning to perceive, and learning routinization of certain classroom processes may be most important. It is a time for acquiring experience for reflection, for sharing that experience, for having someone who helps direct perception to that which is important. But it may not be a good time to bring in instructional theories that have as an unspoken prerequisite classroom experience.

- To furnish the instruction necessary to prepare a teacher for his or her first teaching assignment requires knowledge about what the teacher is going to teach. There is something very odd about having to prepare a beginning teacher for all six elementary grades or all the mathematics curriculum in a high school. It is difficult to become an expert without an area of expertise, such as fourth grade, beginning reading, biology, or government. We should prepare our novices for particular assignments in their first year. Local districts ought to be able to tell new teachers long before school opens what grade or courses they should be prepared to teach. Among the factors that we are sure distinguish the experts from other teachers in our studies was their concern for preparation. They were eloquent

about the need for time to think through what they were going to teach and how they were going to teach it. They put in many hours every week doing this. But in that most vulnerable first year, novices often do not know what they will be teaching, have no lesson scripts or lesson prototypes to rely on, and are given little planning time during the instructional day. Under these conditions it is the rare novice who has a successful first year of teaching.

- Because the goal of many universities is the development of a competent teacher, they may have to redesign their teacher-education programs and take responsibility for their graduates during their first three years on the job, as teachers move from novice to advanced beginner to the competent level. Universities currently turn out novices whose knowledge of pedagogy is limited. The number of required credit hours in education has dropped among elementary-education majors over the last decade, and even fewer hours are accrued by the typical secondary-education majors and those entering fifth-year or post-baccalaureate programs. This reduction in hours has occurred at the same time that our knowledge base about teaching has increased dramatically. The communication of that knowledge base then has to occur in the field, over the first few years, calling for rethinking of the teacher-education curriculum. This is not all bad. It is likely that the front-loading of pedagogical knowledge in the form of preservice course work, a common method of teacher education, was always less effective then desired.

- The novices' relative inexperience in a complex environment allows a good case to be made for the importance of teaching them standard lesson forms and scripts while they are in preservice teacher-education programs. The image of the new teacher as the creative lesson planner, eschewing the teachers' manual and bringing a fresh eye to the creation of curriculum, has already been attacked as unfair by others (Ball and Feiman-Nemser, 1986). It is precisely the new teacher who most needs the manuals, even with all their deficiencies. Teacher educators who try to woo the novice away from inadequate manuals and evaluators who judge the use of manuals in a negative manner should ask themselves where another guide can be found for teaching such things as two-column addition with regrouping for the very first time.

- The teacher-education programs that have tried to make use of the notion of reflective practice or to change the practical arguments of

preservice teachers may be misguided. While skills of reflection — journal keeping, self-analysis of videotapes, intensive discussion with peers of pedagogical issues — are all laudatory goals of a teacher-education program, it must be remembered that the novice teacher may have too little experience to reflect on. From the developmental model I have described, it is more likely that supportive educational interventions to promote reflective teachers have their greatest effect during the advanced-beginner stage of development, and not at the preservice or novice stage. In a similar manner, new programs of preservice instruction that focus on the restructuring of the beliefs and practical arguments of teachers may have also missed the point about the importance of experience and episodic knowledge for a genuine understanding of teaching. Practical arguments end in actions (Fenstermacher, 1986). But until extensive classroom experience has been acquired, there may be too little in the minds of preservice teachers about what actions might be realistic, relevant, appropriate, moral, and so forth. Any analysis of teachers' actions and the practical arguments that they make to justify them must be well contextualized, and thus requires extensive experience.

- The evaluation instruments in use by many states and districts for assessing practicing teachers may not be relevant for those in the proficient and expert stages of development. Many of the evaluation instruments are research-based; that is, they seek to observe, record, and evaluate the behavior that research suggests is associated with effective teaching — teacher questioning patterns, use of advance organizers, monitoring during seatwork, use of waiting time, and so forth. These are desirable behaviors for novices and advanced beginners to learn and to use. But no researcher ever suggested that the causal and the correlational evidence that is useful in identifying teacher behaviors associated with teacher effectiveness was commutative in nature. Great numbers of effective teachers exist who do not do those things at all or do them in some different form than that described by the researchers, as is illustrated by the expert high-school teacher who teaches history brilliantly in a way that does not correspond to the behaviors assessed on the typical teacher-evaluation form (Wineburg and Wilson, 1988). Instrumentation that is perfectly sensible to use with novices may not be suitable for proficient and expert teachers who have found their own ways of accom-

plishing the tasks of teaching. These teachers are more intuitive and holistic in their ways of teaching, and their classroom behavior may look different than that of novices or advanced beginners.

Conclusion

Other policy considerations may be derived from the theory of the acquisition of expertise and supporting empirical data about the differences between experts and novices. Developmental differences are real and they may have important implications for the policies we adopt for training and assessing teachers. Our extensive knowledge base about teaching and assessment should be thought of as more or less appropriate to people in different stages of their development. We can find evidence that preservice education may not be the most appropriate place to teach some things, and therefore we may have to extend our programs of teacher education for some time after our students have gone out into the field. We can find evidence as well that the evaluation of experienced and beginning teachers may have to differ. And there is evidence to suggest that experts, revered as they may be, are not always the best teachers of novices, though they probably are always useful as models.

It is likely that the development of competence out of ignorance, and expertise out of competence, may take a long time in a profession as complicated as teaching. We may not be able to speed the trip up very much because extensive experience is fundamental to development, but we certainly ought to help nurture those willing to undertake the journey by providing the training and evaluation that is appropriate to their level of development.

Funding for some of these studies was provided by the Spencer Foundation, to which I am grateful. Colleagues at the University of Arizona who contributed to these studies include K. Carter, P. B. Clarridge, K. Cushing, J. George, S. Pinnegar, D. Sabers, and P. Stein.

References

Ball, D. L., & Feiman-Nemser, S. (April 1986). Using textbooks and teachers' guides: What beginning elementary teachers learn and what they need to know. Paper presented at the meetings of the American Educational Research Association, San Francisco.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher, 15*, 5-13.

Berliner, D. C. (April 1988). Memory for teaching as a function of expertise. Paper presented at meetings of the American Educational Research Association, New Orleans.

Benner, P. (1984). *From novice to expert*. Reading, MA: Addison-Wesley.

Bloom, B. S. (Ed.). (1985). *Developing talent in young people*. New York, NY: Ballentine.

Bloom, B. (February 1986). Automaticity. *Educational Leadership*, 70-77.

Brooks, D. M., & Hawke, G. (April 1985). Effective and ineffective session opening teacher activity and task structures. Paper presented at the meetings of the American Educational Research Association, Chicago.

Bullough, R. V., Jr. (1989). *First-year teacher*. New York, NY: Teachers College Press.

Bryan, W. L., & Harter, N. (1899). Studies of the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review, 6*, 345-375.

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. C. (1988). Expert-novice differences in perceiving and processing visual information. *Journal of Teacher Education, 39(3)*, 25-31.

Carter, K., Sabers, D., Cushing, K., Pinnegar, P., & Berliner, D. C. (1987). Processing and using information about students: A study of expert, novice and postulant teachers. *Teaching and Teacher Education, 3*, 147-157.

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Clarridge, P. B. (1988). Alternative perspectives for analyzing expert, novice, and postulant teaching. Unpublished doctoral dissertation. Tucson, AZ: University of Arizona.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine*. New York, NY: Free Press.

Fenstermacher, G. D. (1986). Philosophy of research on teaching: Three aspects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.

Housner, L. D., & Griffey, D. C. (1985). Teacher cognition: Differences in planning and interactive decision-making between experienced and inexperienced teachers. *Research Quarterly for Exercise and Sport, 56,* 45-53.

Leinhardt, G., & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78,* 75-95.

Lesgold, A. M. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory.* San Francisco, CA: W. H. Freeman.

Lunn, J. H. (1948). Chick sexing. *American Scientist, 36,* 280-287.

Nelson, K. R. (1988). Thinking processes, management routines and student perceptions of expert and novice physical education teachers. Unpublished doctoral dissertation. Baton Rouge, LA: Louisiana State University.

Peterson, P. L., & Comeaux, M. A. (1987). Teachers' schemata for classroom events: The mental scaffolding of teachers' thinking during classroom instruction. *Teaching and Teacher Education, 3,* 319-331.

Sabers, D., Cushing, K., & Berliner, D. C. (under review). Differences among expert, novice, and postulant teachers in a task characterized by simultaneity, multidimensionality and immediacy.

Schon, D. (1983). *The reflective practitioner.* New York, NY: Basic Books.

Stein, P. & Berliner, D. C. (April 1988). Expert, novice and postulant teachers' thoughts during teaching. Paper presented at the meetings of the American Educational Research Association, New Orleans.

Stein, P., Clarridge, P. B., & Berliner, D. C. (in preparation). Teacher estimation of student knowledge: Accuracy, content and process.

Wineburg S. S., and Wilson, S. M. (1988). Models of wisdom in the teaching of history. *Phi Delta Kappan, 70(1),* 50-58.

A Classroom Teacher's View of the Assessment of Teaching

CLAIRE L. PELTON
San Jose Unified School District

In the final volume of his trilogy on education entitled *American Education: The Metropolitan Experience, 1878-1988*, Lawrence A. Cremin notes that education is "the characteristic mode of American reform." He further claims, "In other countries they stage revolutions. In the United States, we devise new curricula."

Certainly the reform of American education has involved a series of revolutions in my professional life as a teacher and educator. Indeed, since beginning my career as a high-school English teacher in 1955, I have endured four such revolutions in education — approximately one every eight years. The Sputnik curricula of the 1950s, the curriculum of "relevance" of the 1960s, the performance-objectives accountability mode of the 1970s, and the simplistic and naive "back-to-basics" thrust of the 1980s — each of these revolutions that my colleagues and I have *endured* — and I am using that verb deliberately — has had a number of common and unsettling features:

- First, a propensity to initiate these reforms with a great deal of public fanfare, inflated rhetoric, and any number of blue-ribbon commissions on which only a single token teacher — if one was present at all — participated. While I could cite any number of examples, one will suffice. The committee that developed *A Nation At Risk* had one high-school foreign-language teacher among its members.

- Second, the predilection to point an accusing finger at the schools and the educators therein for whatever present ill was befalling the schools. As Arthur Miller once said, "We need a world in which to lay blame." Clearly, it was easier to place blame on the schools than to address the multidimensional and complex issues confronting public education.

- Third, an inclination toward what Larry Cuban of Stanford University calls the "chronic amnesia" that plagues state and national educational reform movements. Such educational revolutions have been largely doomed to neglect or even failure. Having had no voice in the process of reform, classroom teachers have simply closed their doors to the latest educational fad or pronouncement.

- A final characteristic of these revolutions has been an absence of involvement and support in the process of working *with* the schools to find some solutions to the complex problems education faces. Once these blue-ribbon commissions determine what's wrong (usually after one or two visits to a modicum of artfully selected schools), they divorce themselves from further contact or support — financial and/or otherwise.

If I have learned anything through these educational revolutions, it's that we should never discard our old dittos. By way of analogy, let me describe another typical teacher response to these educational diatribes that have been imposed from the outside. Some years back in the early days of the *Berkeley Barb*, some colleagues and I, all clearly members of the establishment, emerged from a College Board meeting in San Francisco only to be confronted by a ragged, barefoot, and luxuriously bearded young man selling a copy of the paper. On the cover was a photograph of a spectacularly posed nude couple clutched together in what can be described only as a pretzel-like embrace. One of the professors took the newspaper and with particular care methodically unfolded the top sheet so the photograph was blatantly visible. He studied it for some time, slowly moving the photograph sideways, then upside down, and then right-side up once more. Finally, folding the paper and handing it back to the clearly astonished young man, the professor shrugged his shoulders, saying, "I've done it."

Similarly, teachers, when confronted by the latest educational contortion, have said to their colleagues, "We've done it." And so in May of 1986, when I first picked up a copy of the Carnegie Report, *A Nation Prepared: Teachers for the 21st Century*, I was also programmed to say, "I've done it." In fact, I was totally unprepared for what I would find in its pages. The majority of the participants on the proposed National Board for Professional Teaching Standards (NBPTS) were to be teachers. As I reread the report, I discovered that an accusing finger of blame was not to be found within its pages. Instead the report

addressed the need to attract quality people into the profession so that we can adequately prepare our students for the world they will face in the 21st century, a world in which communication skills rather than manual skills will be needed. *A Nation Prepared* also underscored the need to address the conditions under which teachers worked, and it promised to assist on a long-term basis with funding as well as personnel. Creating schools that provide a professional environment for teaching, establishing high standards for what teachers need to know and be able to do, and certifying teachers who meet that standard constitute the heart of the report.

So what did I do after reading the report? Like Saul Bellow's Herzog, I sat down to write a letter to one of the names mentioned in the report. In my letter, I described what I have described for you this afternoon — what it is like to have been buffeted back and forth by these educational revolutions, the debilitating frustration that results from the appalling conditions under which teachers have worked, the continuing loss of exemplary teachers most often to industry or to administrative positions in order to earn a living wage, the small but stunning miracles that have gone largely unheralded except by the students who have experienced them.

In the summer of 1986, I was asked to become a member of the Carnegie Planning Committee, whose major responsibility was to put the National Board in place. And in May of 1987 I became a member of the first National Board for Professional Teaching Standards. Since that time, I have crossed the country speaking to several groups about the Board—teachers, administrators, subject-matter organizations, higher-education representatives. In almost every instance, I have first been met with skepticism, but I have left these conferences feeling optimistic that the groups have understood the fundamental differences between the objectives and the function of the NBPTS and those numerous other education-reform commissions that have preceded it.

It is clear that the responsiveness of the teacher audience comes mainly from the fact that a majority of the National Board's members are classroom teachers. Almost equally important, however, is the emphasis in the Carnegie report on the conditions under which teachers are working. Overcrowded classrooms, inadequate supplies, poorly ventilated buildings, schools that are literally in the process of decay, endless supervision duties in the hallways between classes and at lunch, uncaring or uninvolved parents, stacks of meaningless forms to complete, lack of any support or assistance — these intolerable

working conditions have been cited as the primary reasons that an increasing number of teachers are leaving the classroom (Berry, 1988).

Just as serious are the stultifying uniformity and isolation that teachers must tolerate. From 1982 to 1985, I worked part-time as a district collaborator in the Stanford and the Schools Project. Part of my responsibilities was to interview teachers and administrators from San Francisco to San Jose. The most common frustration voiced by teachers was not the salary or the grotesque class size that we endure — frustrations that are clearly valid, by the way — but the isolation and loneliness of the self-contained world in which teachers live. As one teacher said, "I would just like to be able to talk about my teaching with one adult who is knowledgeable in my field. We're all isolated here — locked into the world of teenagers without the perspective of adult vision. I go for weeks at this place without one minute to speak to anyone over 15." No other profession has had to put up with such isolation. And no other profession forces its members to give up the profession in order to earn a decent wage. Physicians and lawyers who advance in their profession are not forced to give up their patients and clients in order to make that advancement. It is an affront and an indignity that teachers have always had to do so.

Lee Shulman's Teacher Assessment Project at Stanford has been working with various teachers from elementary math and language arts, and secondary biology and American history. These teachers respond to the project in a curious but not surprising way that has little to do with the assessment approaches they are developing and evaluating, though they all agree that these are stimulating. Instead, their response has everything to do with the profound difference between the isolation of their usual teaching world and the "brave new world" of teaching they experience through the project. As one teacher said, "This is the first time in my teaching experience that I've had another adult take time to listen to me reflect on my teaching strategies." And another teacher commented, "Meeting these other teachers in the project has made me realize that the problems I've faced in the classroom are not mine alone. I guess that I'm not such a poor teacher after all."

Although the promise of a national certification process for teachers may very well serve to narrow the enormous gaps in both working conditions and in salaries between teaching and other professions, a number of imminent concerns need to be addressed:

- Whatever shape the National Board assessment process eventually takes, it must be both educationally excellent and equitable. The pool of minority teaching candidates has diminished rapidly in recent years. We cannot afford to let this continue. This problem is further exacerbated by the increasing numbers of minorities who drop out of high school. The current rate is 17.5 percent for Black students and 29.3 percent for Hispanic students. The dropout rates in some major cities are even more serious, ranging from 35 percent in New York to as much as 50 percent in Washington (*Business Week*, 9/19/88).

- Changing the working conditions of teachers will cost an enormous amount of money. Federal, state, and local governments, as well as business and labor, must participate in paying for the cost of these improvements. Yet we cannot address new roles for teachers or the restructuring of the teaching profession when classroom teachers lacks sufficient books or even paper and when teaching conditions are unsafe, unsanitary, and unwholesome.

- Between retirement and normal attrition, America must replace one million of its teachers — almost half of the current teaching force — within 12 short years. Just four years from now, we will need at least 200,000 new teachers each year for several years to come. Yet only 8 percent of today's 1.6 million college freshmen demonstrate any interest in teaching, and almost half of those will change their minds before graduation. Traditionally, half of all new teachers leave after seven years. When we have experienced teacher shortages in the past, we have consistently lowered our standards by issuing emergency credentials, by raising class size, or by simply eliminating courses. To continue this "business as usual" attitude toward education may very well portend no business at all for our nation's corporations and for our economic survival. Even so, is an increasingly aging electorate without school-age children going to support the financial commitment necessary to fund education as it must be funded?

- Finally, the National Board will need to move rapidly while the momentum is with us to offer national certification in a number of teaching areas. Our expectation is that national certification will begin in 1993. Yet the processes we are considering are both infinitely complex and time-consuming to administer. On the one hand,

the certification process must be exemplary; on the other hand, we need to move ahead as soon as possible. We have already raised expectations out in the field with the dialogue that has ensued concerning the National Board. Still, when I become impatient, I am reminded of a quotation by one of my favorite philosophers, Mae West, who once said, "Everything worth doing well is worth doing slowly."

Even though the problems I have cited are staggering, they are not insurmountable. And while I have long described myself as basically an idealist trapped in the body of a cynic, I believe that voluntary certification by the NBPTS will offer teachers a level of professional status that is long overdue, one that other professions have always taken for granted. Through a rigorous assessment process, the National Board will certify to parents, students, administrators, and the public at large that Board-certified teachers are among the best and the brightest of all professionals in the nation.

I believe that, with the assistance of a rigorous national certification process, we will see teachers using their considerable leadership skills, both inside and outside the classroom. We will see teachers no longer isolated in a prepubescent or teenage world. Instead, we will see teachers engaged in improving teaching, learning, and curriculum with their colleagues. And we will see teacher-leaders able to earn $60,000 and more as they continue to teach part-time while simultaneously working on staff development, curriculum, and mentoring.

Teaching will become a valued and competitive profession. And the quality of our own lives as well as those that follow us will markedly improve. For the fact remains that the quality of our system of public education, more than any other single factor, dictates the quality of our lives. Our country is just beginning to wake up to this fact. And we must seize the moment. As Samuel Beckett's Didi says to Gogo, "Let us do something while we have the chance! . . . Let us make the most of it before it is too late."

Clearly, certification by the NBPTS will not solve all of education's or society's ills. But it *can* serve as a starting point and a beacon to affirm to ourselves and to the country at large that teaching is indeed a profession — one that carries the same dignity and prestige as engineering, law, and medicine. In short, the teaching profession is finally coming of age and entering the adult world. And I firmly believe that National Board certification will provide the needed public validation and private affirmation of our new status and our longtime worth.

References

Berry, B. (1986). Labor market choices and teacher reform: Policy options for the public schools of the twenty-first century. *Teaching and Teacher Education, 4(1),* 71-81.

Nussbaum, B. (September 19, 1988). Human capital: The decline of America's work force. *Business Week,* p. 114.

Task Force on Teaching as a Profession. (1986). *A nation prepared: Teachers for the 21st century.* Carnegie Forum on Education and the Economy.

New Directions for the Career of Teaching — The Rochester Experiment

ADAM URBANSKI
Rochester Public Schools

Ever since the epochal Massachusetts Schools Law of 1647, free public education has served as the cornerstone of our democratic society. For millions of Americans it has been the principal means for opportunity and upward mobility. Countless parents have made the necessary sacrifices so that their children could get an education and inherit a better life.

For me, as for many others, public education served as a vehicle for opportunity. I came to the United States as one of seven sons of little-educated but determined parents who, in 1957, escaped communist Poland in search of a better and freer life. Nearly four years, a dozen countries, and several languages later, my family arrived in the United States. Public schools helped me build a better life for myself and my family.

My story is not unique. Yet I fear that it is not typical either. For too many students from economically and educationally disadvantaged backgrounds, schools and society in general are failing. Public schools can and should be more effective. But to achieve this, we need major and fundamental reforms. Rather than merely tinkering with the status quo, we must restructure schools, transform the teaching occupation into a genuine profession, and empower all education's stakeholders — teachers, parents, students, and others — with considerably more choices. Public schools of choice can constitute the deregulatory move that permits public education to be more responsive to the needs of a changing society.

Indisputably, the current system is failing. In Rochester, New York, for example, the dropout rate is more than 30 percent; nearly 50 percent of all junior-high-school students fail at least one core academic subject; and only 18 percent earn Regents (academic-track) diplomas.

And lest we assume that it's solely the fault of schools, 80 percent of all children entering kindergarten are already one or more years behind in readiness skills. Many of our students live in poverty, and nearly half come from single-parent families.

While our urban schools are educational "intensive-care units," the nationwide statistics are not more comforting. A recent study of approximately 250,000 high-school seniors by the National Assessment of Educational Progress (NAEP) found that fewer than 35 percent could read at the level of *The New York Times;* only 20 percent could write a persuasive letter; no more than 17 percent could solve a typical two-operation math problem; and a mere five percent were able to read and understand airline or bus schedules (National Assessment, 1986).

A 1986 National Endowment for the Humanities study of what our 17-year-old students know revealed equally depressing findings. One in three, for example, did not know in which half-century Columbus discovered America, or when the Declaration of Independence was signed, or even who the principal adversaries in World War II were. Two out of three didn't know in which half-century the American Civil War was fought (National Endowment, 1986).

Should we then be surprised that 25 percent of all United States college freshmen are enrolled in remedial math classes? Or that 21 percent are in remedial reading and writing courses? There's something wrong with an educational system that yields such lousy results.

When these statistics are segregated along racial/ethnic and socioeconomic lines, we have a virtual indictment of the current educational system. In Rochester public schools, for example, while 20 percent of White students fail a core subject, nearly 65 percent of Black students do so. Success and failure in our public schools are, unfortunately, still inequitable along racial/ethnic and socioeconomic lines.

Common sense tells us that we must change the process of education if we hope to achieve different results. The surest way to guarantee the continuation of the current outcomes is by continuing to do everything the same old way. And while everyone seems to agree that we should support change, there's hardly a consensus on what kind of changes are desirable.

Today's schools were designed nearly a century ago — at a time when the factory system and mass production were the organizational model. Students today are the "product," and teachers are treated more like hired hands than autonomous professionals. This assembly-line approach to educating our students results in too many recalls and

is unsuitable to the changing needs of an increasingly complex society. It promotes the kind of sameness and bland uniformity that makes success and choice — for parents, students, or teachers — difficult at best. If all schools offer the same curriculum, all classrooms are arranged with desks in straight aisles, and all teachers are directed to cover prescribed items in a prescribed manner through prescribed methods, what's left for students or parents to choose? We must first permit — even encourage — different teaching approaches, in recognition of the reality that learning and teaching must be "brain-compatible," and that students have different learning styles and varying levels of intelligence.

In Rochester, we plan to improve the learning environment through a comprehensive effort to restructure schools and reform teaching. The cornerstone of that effort is the Career in Teaching plan — an attempt to transform the single-level teaching occupation into a four-tier profession, as described below:

- INTERN TEACHERS are new practitioners without prior teaching experience. Interns teach under the guidance of more experienced mentor teachers.

- RESIDENT TEACHER status can be earned by teachers who have successfully completed a year of internship but have not yet achieved tenure or received permanent certification to teach.

- PROFESSIONAL TEACHER status is conferred only on those who have earned their tenure and permanent teaching certification — now requiring a master's degree.

- LEAD TEACHERS are selected on a voluntary and competitive basis by a panel that includes other teachers. They teach at least half-time and work also as mentors; or as consultants who select textbooks, write curricula, plan staff-development programs, and direct other instruction-related tasks; or as demonstration teachers who model teaching with an open-door policy. Lead teachers must have at least 10 years of experience, may have to work for up to 11 months, and receive a salary differential. They work with students at risk, teach in remedial and/or enrichment programs, serve as adjunct professors in local teacher-education schools, and perform other duties required of instructional leaders and expert practitioners.

Unlike merit-pay systems that purport to be career-ladder programs, our Career in Teaching plan incorporates the peer-review concept and offers additional professional options to those who qualify. Lead teachers achieve higher status and more pay in exchange for accepting more responsibilities and working a longer school day or year.

The Career in Teaching plan also attacks head-on a major obstacle to effective student learning: the inability to match at-risk students and the toughest teaching assignments with the most experienced and expert teachers.

Under the current structure, the most difficult assignments and the most challenging students often fall, by default, to the least experienced and the most vulnerable teachers. Veteran teachers can choose to avoid such assignments — largely through existing seniority rules. There is probably a correlation between that dynamic and the fact that seven out of every 10 beginning teachers leave the classroom before their tenth year of teaching.

The newly negotiated three-year agreement in Rochester raised starting teachers' pay by 52.4 percent (from $18,983 in 1986-1987 to $28,935 in 1989-1990). Top pay for some lead teachers will be nearly $70,000 in the third year of the contract. The pact also calls for shared governance of instructional programs and school dynamics through a school-based planning process. Teachers will participate in decisions such as filling vacancies for staff positions in their schools. No longer will strict seniority be the determinant for voluntary interschool transfers. Indeed, the school-based planning team at each school will screen all applications, interview all applicants, and select the candidates who best fit the ethos and the particular needs of that school. This form of expanded choice — rather than strict adherence to blanket rules — will go hand-in-hand with the "schools of choice" system district-wide.

The contractual language for school-based planning, perhaps the most promising and the most important of our negotiated "agreements to agree," merely establishes the outside parameters — intentionally omitting the details. The specifics were developed last year, with substantial input from teachers, school administrators, parents, and others. At each school in the Rochester City School District, a team of "stakeholders" will make many of the decisions that, in the past, were made at Central Office or by the school principal alone.

Administrators, parents, teachers, and, at the secondary level, students will make decisions through group consensus. Each school-

based planning team will enjoy an equal voice and will decide on such matters as school budget allocations, school dynamics and procedures, instructional goals, and ways to measure academic progress. The group will even have the ability to seek waivers from school-board policies and from contractual provisions to accommodate joint decisions.

Even more important than the specific provisions of the Rochester contract is the spirit of the settlement. Achieved through a process best described as "principled negotiations," the agreement is based on trust, mutual respect, and labor-management collaboration. Union and management share a joint commitment to the notion that excellence without equity is not worth pursuing; that unionism and professionalism are complementary and not mutually exclusive; that the collective-bargaining process can be used to build a genuine profession for teachers; that all stakeholders in public education should have greater choice; and that teacher empowerment must be accompanied by teacher accountability. But if accountability means assuming responsibility for the decisions and choices that one makes, then teachers, to be held accountable, must not be locked out of the decision-making process.

Admittedly, the jury's still out on the Rochester experiment. If student performance improves, others may decide that investing in teachers may be a model for breaking the cycle of failure — especially in urban public education. But if there is no evidence of appreciable improvement in student learning, then the public may very understandably conclude that they can get lousy results without additional investments — so why throw good money after bad.

The success of the Rochester experiment hinges on whether we will be granted sufficient time to continue the efforts before the clamor for measurable results begins. Too many people, I'm afraid, expect too much too soon. And unfortunately, test scores are still considered to be the preferred indicators of education outcomes.

It's a scenario that repeated itself quite frequently last year.

"All right, Urbanski," someone who stopped me on the street or in a store would say. "It's been months since you got that big fat contract for your teachers. So, are the kids doing better? Are test scores up?"

"Well, real change takes real time," I'd say in an attempt at self-defense.

"That's nice," was a frequent retort, "but you took the money real fast!"

Test scores in our district for the 1986-1987 school year are higher

than last year. Curiously, this should not be cause for celebration — the results probably are not evidence of improved student learning. In fact, students may have paid for the better test scores with a loss of real learning. We must educate the community about the need to change student outcomes in critical thinking and problem solving, as much as we do in traditional areas. Linking program success solely to standardized test scores guarantees continued failure for our students.

We're looking for more effective ways to measure student learning. Meanwhile, the public has a right to ask the question: If not standardized norm-referenced multiple-choice tests, then what?

No one has a satisfactory answer — yet. But the question must be posed to the entire community, and we must launch a search for the answers. The community, through the elected Board of Education, should determine what it is that students ought to know and be able to do at various reasonable intervals. The educators should then devise ways to achieve these goals. And both processes should be perceived as dynamic, flexible, and based on realism.

Accountability need not be perceived as a guillotine: "If you fail, off with your head!" In fact, learning and experimentation cannot thrive in an environment of fear. Instead, we should expect schools and educators to account for and reveal candidly all the information about their enterprise. This includes negative findings as well as positive. Then we should be held accountable for shedding that which doesn't work and building on that which does.

Such a model of accountability would resemble what is now commonplace in the field of medicine. Is a medical researcher, for example, who spends years searching for a potential cure for a type of cancer and then reaches a negative finding fired? Hardly. Instead, the researcher writes it up for the *New England Journal of Medicine* and gives talks about it at medical conferences. Why? As American Federation of Teachers president Albert Shanker says, "So that nobody else dies of the same cure again." And Winston Churchill reminded his citizens once that success is often little more than "going from failure to failure with undiminished enthusiasm."

The challenge for education reform is not merely to shore up a failing system designed for the needs of a prior century. Rather, it's a question of finding new and more effective ways to educate more students. Schools of choice can enhance the other initiatives launched in our district. This can be achieved, eventually, if we trust teachers to design more effective learning environments for their students. That

means doing what we now do differently — not just doing the same things longer or harder. It also means recognizing the fact that the problem with today's schools is not that they are no longer what they once were; the problem with today's schools is that they are precisely what they always were. That's problematic because change is inevitable while growth is optional. Schools cannot afford to fail to grow and meet the demands of a changing milieu.

We are at a pivotal juncture and face a critical choice: Do we constrain ourselves to merely tinkering with the status quo, or are we willing to significantly restructure our schools even if it means taking risks and abandoning traditionally held postures? If we choose the former, we'll continue to get the dismal results that prompted the cry for reform. The latter can offer hope for a much improved teaching and learning milieu.

The education-reform movement heightened our aspirations. And since the most powerful revolution is the revolution of rising expectations, it'll be impossible to unring the bell. Increasingly there will be a willingness to take risks and to try different and better ways to fulfill our mission. Increasingly, we will dwell even more on potential solutions than on past problems. Risking failure is a risk worth taking because so much is at stake.

References

National Assessment for Educational Progress. (1986). *What do our 17-year-olds know?* National Endowment for the Humanities, Washington, D.C.

Gallup, A. M. (September 1986). The 18th annual Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 56-57.

Harris., L. & Associates. (1987). The *Metropolitan Life* survey of the American teacher, 1987: Strengthening links between home and school. *Metropolitan Life, 8*, 56-58.

Perelman, L. (1987). *Technology and transformation of schools*. National School Boards Association.

Raywid, M. A. (February 1984). Family choice arrangements in public schools: A review of the literature. Paper prepared for the National Institute of Education, Educational Policy and Organization Unit.

Restructuring Teacher Education

JUDITH E. LANIER
Michigan State University

My remarks center around two questions: Are the arguments typically made for restructuring teacher preparation reasonable? And what suggestions and justifications for restructuring are getting the most attention these days? Obviously, we should not take the idea of restructuring seriously if compelling reasons cannot be given for fundamental change. Nor should we build new structures without good justification.

Why Should We Restructure Teacher Education in America?

A growing number of educators ground their arguments in the obvious and poignant fact that teacher education in the United States is not well regarded. Not by those who teach, not by those who decide to avoid teaching after their initial preparation, not by the many who leave teaching after only a short service, and not by the veteran teachers who endure. From the first day of autonomous practice, if not before, teachers complain about the inadequacies of their preparation. Continuing and in-service education fare no better. And while criticism is expected from those who pursue professional careers, teachers' overwhelmingly negative, hostile condemnation of their preparation makes teacher education notorious for its ill repute. The fact that teacher education has been disparaged by its recipients for so many decades is a major reason for restructuring. Those with responsibility for teachers' education have a virtual mandate for massive reform.

On the other hand, teachers' criticism of their professional education would not justify restructuring if the overall result of their work were successful. Unfortunately, such is not the case. We are all too

aware of the unflattering portrait of student learning that has emerged from America's schools. Admittedly, part of this troubling condition is a consequence of growing social needs and unmet but rising expectations for more complex learning in an information age. But not all the disappointment is a result of an acceleration in knowledge and skill demands for high-tech living.

Too many students drop out of school because they find it dull and meaningless. Too many measures of student learning produce disappointing results, especially at the level of analytic thought and problem solving. Too many students' attitudes about school are negative, because they associate education with hoop-jumping to get through the system rather than with valuable learning. Such problems characterize learning across virtually all parts of the education system, although they are especially devastating for students from poor schools and neighborhoods. And while we know that multiple factors contribute to the development of children's learning problems, *we cannot pretend that deficiencies in student learning are unrelated to deficiencies in teacher learning. Nor can deficiencies in teacher learning be disassociated from deficiencies in teacher preparation at America's colleges and universities.*

What Are the Major Deficiencies?

Most preparation programs for prospective *elementary teachers* fail because they do not provide depth of understanding in the subjects taught. For all practical purposes, the majority of America's elementary teachers are prohibited from developing the thorough understanding and appreciation for science and mathematics that are needed to teach these subjects with meaning and a sense of passion for ideas. The same is true for other content areas as well.

On the other hand, *secondary teachers* are effectively denied the opportunity to obtain necessary pedagogical knowledge and skill. Most secondary majors enter teaching without knowing how to recognize and help students who lack essential prerequisite knowledge or motivation to learn a given subject.

But the problems are not limited to content-knowledge deficits for elementary teachers and pedagogical-knowledge deficits for secondary teachers. Both elementary and secondary teachers are denied the opportunity to obtain significant knowledge about cultures differ-

ent from their own. Critically important understandings that enhance teaching diverse youngsters effectively in a multicultural, increasingly international society are left to trial and error and on-the-job learning — experiential lessons that are often counterproductive. The majority of teacher-preparation programs provide little substantial opportunity to learn about teaching children with special needs. Most elementary and secondary teachers do not have a chance to learn about the educational organizations and communities in which they will work. They have yet to develop the knowledge and skills that permit effective participation as members of educational teams. They lack the analytic processes that help teachers make increasingly effective decisions under conditions of growing complexity and uncertainty. And most would-be teachers learn little, if anything, about the use of technology to improve learning — an obvious oversight in our high-tech society.

Almost all prospective teachers are denied the opportunity to learn the practice of teaching gradually, over time, in a range of diverse instructional settings. Few instructional programs permit serious examination of pedagogical content knowledge. And the chances are very small that prospective teachers will see demonstrations of, or have direct experience with, content-specific pedagogy that emphasizes higher-order thinking for *all* students in more or less standard classrooms. Far too many prospective teachers are themselves subjected to bad instruction, useless course work, and poor clinical experience and supervision. Curriculum coherence in the professional program of study is usually questionable, if not absent, as is the investment in any serious accreditation efforts that would strengthen quality assurance.

There are many sins of omission and commission in American teacher education, but the major reasons for restructuring are:

- Practicing teachers express continued dissatisfaction.
- Students are not learning as much or as well as they can and must.
- Our society's economic and political need for more knowledgeable, skillful, and successful teachers (and learners) continues to mount, as we identify more and more inadequacies in the preparation programs afforded prospective professionals.
- Clearly, we *can* and *must* do better if substantial numbers of teachers are to be successful in their work and in their efforts to become certified by the National Board for Professional Teaching Standards.

What Kinds of Restructuring Are Called For?

It seems fair to suggest that a major overhaul is in order — in contrast to continued tinkering with a few courses in the more-or-less standard programs that characterize teacher education. America's professional education for teachers needs *sweeping change* that will help us achieve an entirely new level of quality. Such a jump will enhance teachers' success in their work, permitting, in turn, increased rates and levels of school learning heretofore unrealized. But a major restructuring of teacher education must address emerging social changes, including changes in the nature of teaching, the shifting career patterns of teachers, differences in learning expectations for American citizens, and constantly evolving student demographics.

The teacher-preparation programs that are now so sharply criticized were designed for the occupation of teaching that evolved in America more than a century ago. The occupational patterns and dynamics that shaped the early character of teaching are now out of date, although their imprint remains in the institutional arrangements for educating teachers. To think about the nature of the needed restructuring, we must understand these occupational traditions as well as their influence on today's dysfunctional programs.

Traditions in Need of Change

Teacher preparation has been a mass-production enterprise — not only because America has needed many people to teach its many students, but also because most teachers haven't stayed in the occupation very long. Replacement troops must be recruited on a regular basis. This very large, high-turnover teaching force has created conditions similar to those associated with transient labor for other kinds of work. Consider the following common characteristics:

- Modest investment in education and training
- High expectations for the performance of physical work and the exercise of simple, routine tasks
- Low expectations and restricted opportunities for intellectual analysis and judgment
- Expectations that employees should do as they're told rather than participate in management decisions

- Modest selectivity for the job, since most anyone can do the work at an acceptable level
- A work force comprised of many young people and a large proportion of women and minorities
- Low pay and few occupational perquisites
- Low status overall, combined with a general flight from the occupation whenever better opportunities present themselves

These characteristics have also affected teacher education. Any college can do it, and most any program will suffice. Selectivity of students for programs is modest at best. Getting in and getting out are easy, as few candidates are denied either entrance or exit. The programs are low-investment items, which contributes to the perception that the practice is not particularly complex or important. The content is neither significant nor deep, since teachers will merely do as they are told anyway. Learning to teach can be done, and is done, on the cheap. A costly education would be impractical, since the young recruits are not apt to stay.

The professional unit that houses the teacher-education program is invariably among the lowest-status units at a college or university. The faculty responsible for the program are often part-time and temporary, with little authority or control. Just as there is flight from the occupation of teaching in the schools, there is faculty flight from teacher education in colleges and universities. Teacher education is seen as low-status work, done by less-than-distinguished faculty who help undistinguished students learn weak and technical curricula under a set of program conditions and requirements that lack coherence, rigor, and currency.

The social and academic circumstances that foster these legitimate stereotypes must be recognized and understood. And the teaching profession, together with the professional schools that share responsibility for the initial and continuing education of teachers and other school professionals (e.g., psychologists, counselors, administrators, special-education consultants, etc.), must work collaboratively to change them. Change is necessary if we are to adjust appropriately to the occupational features, professional practices, and institutional settings needed for teachers in today's postindustrial, new intellectual era.

Rising Expectations and the Need for Change

Expectations for America's public schools continue to climb as calls for reform emanate from virtually all sectors of society, including professional, government, business, and citizen groups. The calls are unanimous in their demand for fundamental change: present dropout rates can no longer be tolerated; students' test scores, especially those indicative of problem-solving and analytic capacity, must rise; all students must acquire scientific and mathematical literacy; and schools must become more responsible and accountable for such learning. Yet most everyone recognizes that changes of this magnitude cannot be brought about without a much greater proportion of well-educated, caring teachers than we have today.

Restructuring for Stronger and More Professional Programs for Teachers

Some key elements of the needed restructuring are obvious; others are more subtle. First, the insidious form of tracking that now captures most would-be teachers in the baccalaureate years must be recognized for what it is: a less-than-adequate collegiate education for the 21st century. Prospective elementary teachers, in particular, are often placed in watered-down, segregated classes (e.g., math for teachers, science for teachers) and given a more limited and less rigorous education in the disciplines than their supposedly "more intelligent" counterparts in other major fields of study. These stifling educational practices must end. Prospective teachers, *together with* their preprofessional counterparts who are pursuing other important careers (e.g., engineering, medicine, law, public policy), must experience higher-quality undergraduate studies. The "teaching as telling, learning as accumulating, and knowledge as facts" traditions need to be changed at the baccalaureate level if we want to change them in elementary and secondary schools. If the major and minor areas of undergraduate study are not more liberating and better connected to meaningful learning and living in our contemporary world, education will not improve much in our elementary and secondary schools. Teachers cannot develop in their students what they themselves have not learned and experienced intellectually.

So the Holmes Group and many other reform-minded educators have urged an end to undergraduate majors in education and suggested replacing them with improved content and pedagogy in selected disciplines. This means *different* and *better* disciplinary majors than are currently available in most baccalaureate programs. Many of America's leading universities are replacing the undergraduate education major with content and pedagogy in selected disciplines designed to provide all students — not just prospective teachers — with a mature understanding of the knowledge that constitutes school subject matter. To complement improved and strengthened disciplinary learning in the baccalaureate years, Holmes Group institutions are also developing preprofessional learning experiences in education. Such courses and experiences, more or less equivalent to a minor, are meant to perform several functions. As part of a university's general-education program, they can help to recruit undergraduates to the study of teaching by exposing them to a variety of engaging issues and providing early field experiences in education. Students who are not attracted to a teaching career can benefit from these experiences by bringing a more sophisticated understanding of educational issues and questions to their lives as citizens (i.e., as school-board members, parents, and informed voters in their communities). Preeducation minors can also nurture the interest of undergraduates aspiring to teaching careers and begin to prepare them for professional study in education. Together, these strands of the undergraduate curriculum can deepen a prospective secondary teacher's disciplinary content knowledge and pedagogical understanding of a subject. In addition to providing an opportunity for elementary teachers to master a limited range of disciplines, the preprofessional program can inspire them to collaborate with their colleagues to round out their professional repertoires in other subjects.

In addition to reworking the undergraduate, preprofessional curriculum, colleges and universities must overhaul the professional-studies component of the program. Moving the bulk of the professional certification studies to the postbaccalaureate level would allow teacher-education programs to be planned and articulated more coherently. A graduate-level professional curriculum would include foundational as well as pedagogical studies and connect them more closely with extended multisite field experiences, internships, and residencies. In general, an integrated curriculum would better ensure that graduates are able to provide the highest levels of subject-matter learning for

all students. Over time, prospective teachers could develop and refine the ability to construct learning communities in their classrooms and schools that engage children and adults alike in the most rewarding forms of problem solving. And they could develop a commitment to professional inquiry and norms that steadily improves their practice.

Reconceptualizing the entire undergraduate and graduate curriculum in this way does not increase the educational investments of most teachers radically, since a majority of them now routinely obtain master's degrees within several years after initial licensure. Through the creation of more integrated, coherent experiences, the extended course of study can respond to evolving norms and expectations for student learning and teacher practice. It can also complement contemporary patterns of teacher employment. For the first time, teaching has become a career rather than a principally transient job. An investment of five or more years in professional preparation and development would have been unreasonable to expect, and probably impossible to get, when teaching careers lasted no more than two or three years. Men and women who are willing and able to make teaching a career should be willing to invest modestly in the kind of education and experience that prepare them to practice in their classrooms and assume broader responsibilities as community and public leaders. Traditional undergraduate programs do not include sufficient time to master subject-matter content and related pedagogy, to develop the appropriate ethos for enriched professional roles and responsibilities, and to participate in long-term supervised residencies.

Second, the labor market for teachers has changed in unprecedented ways. Teaching no longer can count on continuing access to a segment of the labor force with few occupational alternatives, a change that promises to alter teaching significantly. Competing with other desirable occupations for its fair share of talented individuals obligates teaching to cast its net broadly, without compromising professional standards for knowledge, skill, and character (as communities have been tempted to do on a number of occasions in the past). Trustworthy alternative routes into teaching must exist in order to expand the talent pool to the necessary dimensions. Liberal studies in education at the undergraduate level will help to nurture and deepen the commitment of students already interested in teaching careers and help recruit to teaching others who at age 19 or 20 are flirting with a variety of career possibilities. Moving certification programs to the graduate level will help to open the profession to individuals who decide somewhat later

in their education to prepare for a career in teaching. Offering alternatives for both full- and part-time students will further broaden the applicant pool and increase the flexibility of entry to teaching.

Third, although occasional courses might have prepared prospective teachers for working in isolated classrooms in egg-crate schools, most teachers need a collegial ethos at work. This type of atmosphere requires a professional education and induction experience organized around faculty and student cohorts, which encourages faculty members to become deeply engaged in the academic and professional lives of their students, to work with them over an extended period of time, to become conscious of the program's cumulative impact on students, and to minimize the bad effects of piecemeal course development. Cohort arrangements increase the level and strengthen the quality of faculty participation by developing the mentoring relationship to the fullest. Organizing program participants into reasonably stable groups will engage faculty members positively and improve the educational experience of students considerably. Instead of wandering through courses in isolation, preparing for a life of solitude in schools, with class choice largely a function of schedule convenience or rumors about the level of class difficulty, prospective teachers will complete a coherently developed program that celebrates achievements and provides opportunities to struggle with problems in collaboration with colleagues.

Linda Darling-Hammond captured the traditional relationship of teachers to their professional knowledge memorably: "Figure it out for yourself, do it all yourself, and keep it to yourself." This pattern aggravates already difficult problems of practice and leadership in schools; it must be broken. Teachers are becoming aware that isolation exacerbates many of their difficulties. Like other workers today, they are realizing that genuine solutions require collaboration and collegial support. No one can know everything, but collaboration with committed colleagues provides access to a variety of powerful complementary skills. Prospective teachers will expect to work in supportive collegial environments if they attend institutions that expect that much of them.

Fourth, traditional student-teaching practices, abbreviated and ordinarily restricted to work with a single cooperating teacher, cannot provide prospective teachers with the breadth of clinical experiences that professional practice requires. Restructuring teacher education would provide opportunities for extended internships of an academic year or more in length that could support experimentation and the integration

of subject matter and pedagogical learning far better than is possible in today's hurried field exercises. And prospective teachers must learn to anticipate — and be comfortable — working with diverse learners. Multiple-site internships or assignments in schools serving substantial numbers of learners who have not been well served by traditional classrooms can provide this sort of experience.

Fifth, it is unlikely that tomorrow's teachers will acquire the learning they need in today's schools. This is why Holmes Group universities are working with surrounding districts to create "professional-development schools." These institutions, analogues of the teaching hospital in medicine, will be staffed cooperatively by district and university faculty and prospective teachers who are on site for various field experiences or extended internships. Professional-development schools will exemplify the best in practice and serve as centers of inquiry about effective policy and methodology.

Professional-development schools must emerge within the context of long-term partnerships between universities and school districts, with the goal being mutual change in both. These arrangements cannot be short-term interventions by universities in the lives of children and teachers. They must be enduring new organizational plans devoted to mutual change. For the partners, as well as for their different students, professional-development schools will become communities of learning that blend research and practice to the mutual benefit of the schools and universities. In a sense, these schools will be centers of applied research, testing research understanding under real-world conditions and using research to change our sense of what is achievable in the real world of schooling.

Sixth, professional-development schools and other restructured institutions will expect and encourage teachers to assume a variety of new responsibilities, both in the classroom and out. While not abandoning the classroom, some teachers will contribute to vital work in curriculum improvement, evaluation, and the education and induction of novice and continuing teachers. They might also collaborate in integrating school learning with carefully considered community-development projects. Enriching the professional role of teachers will minimize the damaging effects of a flat career structure, help to halt the defection of ambitious and energetic teachers from schools, and allow schools to use thoughtful, skilled employees to the greatest extent possible.

Seventh, the delegation of responsibility and authority for the education of teachers needs to change. Up to this point, the discussion about restructuring teacher education has focused principally on teacher-education institutions. Prospective teachers, however, receive most of their education not in professional-study programs but in departments of arts and sciences. Although teacher-education programs are ordinarily held responsible for the education of teachers, they typically provide a small percentage of each prospective teacher's relevant course work. It is customary, for example, for prospective secondary teachers to take only three teacher education courses from a school or department of education — roughly eight percent of their entire bachelor's degree program. The university's general-education division and academic subject-matter departments offer almost 85 percent of the remaining courses, with another seven or eight percent of the credits acquired in student teaching — an experience that is ordinarily handled by a cooperating high school. Even prospective elementary teachers who major in education take only one-quarter to one-third of their credits in their schools, colleges, or departments of education.

Heretofore the academic departments — in which most of the education of teachers occurs — have not been held responsible for ensuring that prospective teachers grasp subject matter with the genuine understanding and maturity that would enable them to teach it to others. We continue to see, in the memorable phrase of David Clark and Gerald Marker, "the assignment of responsibility without authority and authority without responsibility." An integrated undergraduate-graduate model will help to fix responsibility and authority by bringing the academic disciplines to the foreground at the baccalaureate level while keeping professional studies in the background. It will also bring professional education to the foreground at the masters' level while placing disciplinary learning in the background.

Restructuring the education of teachers — not just restructuring teacher education — must engage faculty from all disciplines. Until now, discussions about knowledge have occurred in isolation, with education faculty focused primarily on professional subjects, and faculty from the academic departments concerned with what their graduate students need to know. Restructuring the entire enterprise requires both independence and collaboration among faculty in professional-education programs and faculty from the disciplines. For prospective teachers, and probably for those majoring in other disciplines, disci-

plinary knowledge must be encountered in a fashion that leads to understanding and the ability to work with the knowledge maturely, to create, for example, or to visualize and articulate connections, relationships, and applications. This kind of learning can be achieved with far more efficiency and effectiveness through courses of study that are more coherent than those that undergraduates encounter today. Restructuring the education of teachers requires a new thoughtfulness about the issues of responsibility and authority that Clark and Marker (1975) raised more than a dozen years ago, and a willingness to collaborate wherever appropriate — and to continue to work independently as well — to ensure that all students learn something worth teaching well enough to teach it.

Conclusion

In light of the immensity of the task of restructuring teacher education, it is reasonable to ask whether or not other approaches might accomplish the same ends more easily or more predictably. Anyone who knows anything about the difficulty of restructuring institutions of higher education and licensure can legitimately inquire about the possibility of circumventing the entire task altogether, by deregulating entry into teaching, for example, or by turning exclusively to the enterprise of assessment in order to find and reward the "best" teachers.

As tempting as it might be to endorse alternatives to restructuring teacher education, there are compelling reasons to strengthen the tripartite standard of comprehensive controls — professional education; extended, multisite supervised residencies; and thoughtful assessments — on entry into teaching.

Teaching is in many respects a unique occupation. We should be more cautious and careful than other professions about the preparation, licensing, and placement of our practitioners. The children who are the primary recipients of teaching's professional services are distinctly vulnerable. Until a certain age, they have no real choice in the matter, since schooling is compulsory. They are not consenting adults, and therefore deserve more, rather than less, protection from inferior practice than do the clients of other professions. They have little power or authority to recognize, influence, resist, or express dissatisfaction with poor practice, except at enormous personal cost by withdrawing or dropping out of school.

Furthermore, the standard for preparing and entering teaching should be conservative, because schooling is a public enterprise. We all must pay for our schools regardless of whether they employ competent, trustworthy teachers. Because schooling is both publicly funded and compulsory, it is essential to minimize the risk of permitting imposters and quacks to practice as autonomous professionals. While we may wince at the thought of imposters and quacks in teaching, we all know that "competent" teaching is easy to fake, since "keeping classes" or "dispensing information" are too readily confused with real teaching.

Prevailing arrangements in most schools also make it difficult to remove imposters who find their way into professional teaching roles. Other occupations are able to rely on market mechanisms to expose and drive out ineffective or dangerous practitioners. Again, teaching is unique in this regard. Traditions of work in relatively unsupervised, isolated classrooms housed in essentially monopolistic institutions make it difficult to do anything more than transfer weak teachers — hoping that they will improve on their own in another environment, while fearing that they will simply find another place to hide. If errors are made, and errors will be made even when comprehensive checks are in place, there is no easy or convenient means of correction. It is thus our moral obligation in teaching, more so than in other professions, to do whatever we can to avoid betraying the trust that our children and our taxpayers have in our schools and their teachers.

We must, therefore, design and enforce the strongest checks on practitioner quality. Can we be any less demanding than the strictest professions in making every effort to ensure that only highly qualified teachers are granted the privilege of professional autonomy? In order to minimize the risks of incompetent practice, the medical community established a conservative standard of professional education and induction parallel to the one inherent in the restructured model of teacher education that I proposed earlier. Can education afford to be more reckless than medicine by choosing to rely on something less than the tripartite standard? Children can be maimed in mind and spirit just as they can be maimed physically.

Restructuring the enterprise will take much time, and it will be painful. And although considering alternatives is vital and will contribute much to achieving the goals of strengthened teaching and learning, it is essential to resist the temptation to make deals that seem too good to be true, for they inevitably are.

The Case for a Supervised Teaching Internship

LINDA DARLING-HAMMOND
The RAND Corporation

All professions face a quandary when they seek to establish a certification system that will assure the public of new members' fitness to practice. The best assurance is provided by allowing the novice to practice and then determining that he or she is practicing competently. However, the demands for reliability and validity in that determination, along with the need to ensure the safety of prospective clients, have obliged professions to create a series of assessments for licensure. Although each of these assessments weighs reliability, validity, and job relevance differently, the series, taken as a whole, convinces the profession and the public that the entrant is fit to practice (Wise and Darling-Hammond, 1988).

Most professions have adopted requirements for a clinical internship or apprenticeship experience as part of this determination. There are two reasons. The first is that the candidate requires an internship to learn how to apply knowledge and make decisions appropriately. The second is that examinations — even performance tests — are incomplete tests of the ability to use knowledge and apply skills. By requiring the candidate to successfully complete an internship, the profession has another measure of performance that is valid on its face, because it involves the actual delivery of service to clients. But because the rating of the intern's performance is necessarily subjective and unstandardized, this measure cannot be used as the sole basis for the licensing decision. The dual requirement of satisfactory completion of an internship and an examination balances the demands for reliability, validity, and job relevance. It also provides the assurance that neither the internship nor the examination alone can provide.

As is true in other professions, existing instruments and methods for assessing prospective teachers do not fully capture the complexities of teaching knowledge and the context-dependent nature of teaching judgment. And because the acquisition of teaching skill is so dependent on developing judgment in complex, nonroutine situations, this skill cannot be adequately assessed until after the candidate has encountered and worked through many of the common problems of teaching practice.

Before new teachers are granted a continuing license, an evaluation process should identify whether or not they possess a range of needed teaching skills. These skills cannot be acquired through formal teacher education alone. There is also no assurance that they can they be effectively acquired by trial and error during the initial years of practice. If a major goal of licensure is to increase the probability that those admitted to practice can make appropriate decisions and execute sound teaching strategies effectively, the licensure process should require both *evidence of* and *opportunities for* learning these skills.

Other professions, such as medicine, architecture, psychology, and engineering, have addressed similar concerns about the development of skill in practice by requiring a form of structured internship prior to licensure. This internship serves simultaneously as a training vehicle, a safeguard for the public, and a source of support and assistance to beginning practitioners. The internship experience extends professional training in a clinical setting without exposing clients to unsupervised novices or leaving to chance the acquisition of essential skills. A set of educational goals specified by the profession is incorporated into the design of the internship program. Not incidentally, this type of program is more oriented to assistance than to assessment, and to enhancing effectiveness than to screening and sorting. These latter functions are left largely to the licensure examination, thus untangling the formative and summative aspects of the evaluation process.

In contrast, beginning teachers — with no further assistance from their college professors and few school-district resources available for formal support — are generally left to sink or swim during their first years of teaching (Wise, Darling-Hammond, and Berry, 1987). Research on the experiences of beginning teachers confirms that the likelihood of long-term success is substantially impaired for many by the absence of expert guidance, support, and opportunities to reflect on their efforts (Ryan, 1980; Tisher, 1978; McDonald, 1980). These initial teaching experiences have far-reaching effects, for "the conditions under

which a person carries out the first year of teaching have a strong influence on the level of effectiveness which that teacher is able to achieve and sustain over the years, on the attitudes which govern teacher behavior over even a forty-year career, and indeed, on the decision whether or not to continue in the teaching profession" (National Institute of Education, 1979).

The Importance of Teaching Knowledge

In her presentation on teacher education, Judy Lanier made a very important point: *Teacher learning matters*. This idea has not always been incorporated in thinking about how to improve education. This conference and the education-reform movement's efforts to professionalize teaching aim to increase our understanding of *how* it is that teacher learning matters, and thus to legitimize investment in teacher learning.

Preparing knowledgeable teachers has not typically been the mode of investment chosen for education reform in this country. Many approaches to changing the quality of education have assumed that the teacher is a conduit for policies, curriculum packages, and rules, regulations, and mandates of various kinds. But they have not assumed that what the teacher carries in his or her mind is a very important part of what happens in the classroom. Consequently, our educational system has underinvested in teacher education; it has invested not at all in teacher induction in the initial years on the job; and it has invested very little in ongoing professional development. Until now, an insufficient commitment to the importance of teacher learning has kept teaching from becoming a profession.

Because the knowledge that professionals hold defines them as professionals rather than as bureaucrats or factory workers, teacher knowledge is crucial. The nature of learning and of effective teaching demand informed decision making. If students all learned in exactly the same way, at exactly the same rate, in predictable fashion, and if they came to school as standardized as pieces of metal on an assembly line, we would not need to worry about teacher learning or the professionalization of teaching. If students were uniform in their needs, learning styles, preconceptions, dispositions, and stages of development, teachers could use a simple cookbook approach — step 1, do this; step 2, do that; step 3, do the other thing — and the outcomes would be certain. We know, however, that students learn differently.

These differences pose complex and varying challenges for the teacher that have important implications for how effective teaching should be evaluated (Darling-Hammond, Wise, and Pease, 1983). Nonetheless, most schools — and much of the surrounding policy-making framework — are structured around assumptions that students are all alike. So these new proposals for assessing teacher learning involve a revolution in our thinking about how learning occurs as well as what teaching is.

Three aspects of socialization to teaching have defined the approach of schools to teacher learning in the past. If we could encapsulate these as proverbs for beginning teachers, they would be, "Figure it out yourself," "Do it all yourself," and "Keep it to yourself."

Even the best pedagogical preparation in a school of education can go only partway toward the acquisition of clinical skill. When it comes to that very important and very difficult kind of teacher learning associated with applying knowledge to the many problems of teaching practice, schools have said, "Figure it out yourself." Furthermore, beginning teachers are usually placed in the least advantageous situations, with the students no one else wants to teach, in the schools with the highest turnover (because many teachers with seniority have transferred out). Consequently, beginning teachers often find themselves in schools in which the potential mentors are fewest (Wise, Darling-Hammond, and Berry, 1987). Someone hands the beginner a key to a practically empty book room at the beginning of school and says, "Figure it out yourself. We'll see you in June . . . if you make it that long!"

Second, "Do it all yourself." Teaching has not yet developed the mores of consultation and collegiality that encourage colleagues to confer about problems of practice. In highly developed professions such as medicine, it is considered useful for a physician to consult with another doctor about a patient. In teaching, it is generally viewed as a stigma to ask someone else for advice or to venture a suggestion to another teacher.

That goes along with the third feature, "Keep it to yourself." Because of the closed-door ethic in many schools, it is a dangerous activity to proffer unsolicited advice to another colleague. Many educators seem to believe that all knowledge should reside and be kept in one's own head. And unfortunately, some of the incentive systems currently being urged on schools in an effort to apply "business principles" to education only encourage competition among teachers and

reinforce some of the existing dysfunctional features of teaching. Colleagueship and consensus about standards of practice are fundamental objectives that reforms in teacher preparation, including the idea of a supervised internship or residency, are beginning to try to address.

The Evolution of Professions

Over the better part of the last century, debates have been conducted in many professions about, first, whether training is necessary for practice, and second, what the appropriate forms of preparation and training ought to be. These debates took place in medicine at the end of the 19th century. There was a large and vociferous debate about whether one needed any training at all to be a doctor. Some said — just as some now say about teaching — that people were born to heal, not taught to be healers. These debates occurred in the legal profession in the 1920s, and in architecture and engineering more recently. Every occupation that is now regarded as a knowledge-based profession has had to go through the process that teaching is going through now. In each case, the occupation had to address two questions: "Is there something special that practitioners need to know?" and if so, "What is it?"

In virtually all of these other professions, after the initial debates about whether training should be required were resolved, the occupations began to move from a model of the self-taught practitioner to an apprenticeship approach. So doctors followed other doctors around in the buggy to learn to become physicians. Lawyers engaged in clerkships, architects pursued apprenticeships, and so on. As practitioners developed more formal knowledge and discovered the limits of modeling for efficiently conveying large bodies of knowledge, the occupations began to require more formal preparation. Sometimes formal education substituted for the apprenticeship (or vice versa) in a hiring or certification decision. In some of these occupations, the apprenticeship almost disappeared as formal preparation took over as a way of readying practitioners. But gradually in all of them, the need for both book learning and practical training and the contributions of each to the development of expertise were acknowledged.

The major realization these professions have reached is that the two

forms of training — didactic and clinical — are different and complementary parts of the equation of learning to practice. Also, the role of the internship as a way of evaluating competence — along with evidence from coursework and examinations — has been established in all of these professions. So, as Lee Shulman suggested this morning, a union of complements has been forged in assessing the skills that are sought in these occupations.

Structuring Professional Preparation and Assessment

A number of other professions have structured their preparation and assessment for licensure in similar ways. Generally, formal university-based education (often followed by a test of basic knowledge) is completed prior to a structured internship. Successful completion of the internship is a prerequisite to sitting for a test of clinical skills. In engineering, for example, a candidate completes an undergraduate engineering degree and then a test on the fundamentals of engineering. After four years of professional experience, the candidate submits detailed summaries and work samples to the state board for evaluation. Then the candidate can sit for an examination on the principles and practices of engineering. The clinical skills test can only be taken after the apprenticeship has been completed. Character references are also required in this and other professions as part of the ethical requirement for certifying a professional.

Licensed psychologists complete a doctoral degree and an internship program that includes a one-year predoctoral and a one-year postdoctoral clinical experience. The institution granting the Ph.D. verifies successful completion of these internships. Then the candidate may sit for an examination on practice in psychology. In most of the medical-specialty board fields, candidates must take a test in basic medical science and a test in the clinical sciences after completing a graduate medical degree. Then they undertake a two- to five-year residency program, followed by written and oral examinations of their clinical skills.

In terms of professional-education requirements, architecture has reached a stage very similar to that which teaching is now entering as it strives to extend the time allotted for professional education. An architecture candidate may complete any one of three courses of study:

- A five-year program in architecture,
- A four-year bachelor's degree in architecture followed by a two-year master's degree program in architecture, or
- A four-year liberal arts degree and a three-and-one-half-year master's degree program in architecture.

But then all candidates must complete a two-year structured apprenticeship in an architectural firm, where they accrue a specified number of hours of practice and observation in 14 task areas. The candidate is supervised by an outside educational adviser and a sponsor inside the firm who must certify successful completion before the candidate can sit for the Architectural Registration Exam. The examination tests knowledge in a written format as well as design ability in a performance assessment.

What you've heard today about the ways in which teacher assessment is evolving mirrors in many important respects the progress that other professions have already made. In some cases this progress is quite recent: The structured internship program in architecture was not adopted until the late 1970s. An earlier unstructured apprenticeship requirement (rather like the mentoring programs recently begun in teaching) was replaced, because professional architects felt that enforcing standards necessitated creating an apprenticeship that could define professional practice and ensure exposure to the required knowledge base.

Defining the content of an internship is also important for ensuring fairness and equity in licensure. If we want tests of teaching performance to be fair and reliable for all candidates, it is absolutely critical that the internship program be structured so that everyone has an equal chance of being exposed to the concepts that will be presented. Most current induction programs for teachers do not attempt to define what goes on between a mentor and his or her pupil. They also do not ensure that the novice is exposed to a range of teaching experiences. Consequently, the induction experience, while it can be very valuable and is certainly better than nothing at all, is an idiosyncratic one. It does not systematically represent the kinds of knowledge and skills we want to be sure that people accumulate in the process of learning to practice.

Standards for Teaching

The importance of establishing standards for entry to the teaching profession has been fully discussed in the policy arena and the professional community and more or less accepted in the last 10 years. But standards do much more than determine who will and will not teach. They define the nature of teaching desired and the effective teacher. The most important function of professional examinations is defining the knowledge and skills required of practicing professionals. The tests, including the requirements for a residency or internship, actually are the means by which the profession makes an explicit statement about what is worth knowing and how it should be known and demonstrated. So the content of these instruments is extremely important for the preparation of professionals.

The statements encapsulated in the requirements for internships and examinations exert a powerful influence on training and practice that is totally independent of cutoff scores or passing rates in a given year. Of course, these figures are important concerns. But only recently have educators begun to attend seriously to the *content* of the examinations, which is by far the more important part of this activity for the profession and, ultimately, for the education of children. If teacher learning has something to do with student learning, then it is very important that what teachers learn be useful for the encouragement of student learning.

Standards, then, focus in two directions. They reflect backward to teacher preparation and forward to teaching practice. But they are not generally operational statements. They have to be translated into preparation programs on the one hand and into testing requirements on the other.

These translation activities are critical. Frequently in education, problems begin at this stage. Excellent statements of standards can be trivialized when goals become narrowly framed learning objectives in curriculum guides, or when standards for good teaching are reduced to observable behaviors in evaluation instruments. While there is nationwide interest in testing teachers for entry into the profession, there has not been until recently a corresponding level of interest in the process of translating standards into effective programs for teacher preparation and evaluation. Now, fortunately, the attention of

researchers and practitioners is focused increasingly on the improvement of beginning-teacher evaluation.

A licensure process should ultimately measure several different kinds of knowledge and skill. As we heard this morning, the knowledge that underlies good teaching is not always obvious from the activities you might observe when you watch good teachers teach. Nonetheless, that knowledge is important to effective teaching, and there are important reasons to assess the acquisition of that knowledge base directly, along with measuring skill attainment. This is a critical point for improving teaching because much teacher preparation and induction has been solely technique-oriented. Often teachers are trained in specific methods without sufficient grounding in the sciences and theory that underlie judgments about appropriate applications. If we were making an analogy to medicine, it would be like training doctors in techniques of treating heart disease without giving them the background in pathology, physiology, anatomy, and epidemiology that permits diagnosis and evaluation of alternative treatments. Teachers also need an understanding of the fundamental sciences that underlie pedagogy: human behavior and motivation, cognitive psychology, child growth and development, and so on. Teachers who acquire specific techniques without that foundation may find that when a particular tactic fails to work, they have no way of evaluating why or of deciding what option should be tried next. Because one cannot infer the mastery of a broader base of knowledge from the exhibition of a specific technique, both training and assessment need to incorporate and evaluate such knowledge explicitly.

Obviously, we also want to encourage candidates to acquire the ability to apply knowledge appropriately in different contexts and to use skills in performing diverse teaching tasks. Inferring performance skills from knowledge is also not useful, because a candidate who knows a great deal about teaching is not necessarily able to handle the thousands of things that are happening simultaneously in a classroom. Learning to manage 30 children's different personalities and needs, while prioritizing and juggling often-conflicting goals, doesn't happen quickly, automatically, or easily. These are skills that have to be developed. They can be taught, and that is the rationale for a teaching residency. Such an approach will acknowledge the complexity of learning to teach and increase the probability that beginning teachers will learn not just to cope, but to teach effectively.

The Design of a Teaching Internship

In teacher education, students should learn the sciences basic to teaching and the methods of effective teaching; in the internship, they should learn to apply that knowledge to teaching situations. Skills in managing the learning process are developed by observing and working with students and by selecting learning materials, assessment tools, and teaching strategies under the guidance of expert practitioners. As interns progressively gain more knowledge and skill, they should be given greater opportunity to make decisions and teach students, but always under supervision.

To achieve its goals, the internship must be structured carefully to provide supervised clinical experiences that cover all major domains of teaching practice and include opportunities to learn about the variability in students' learning styles and stages of cognitive development. The internship must also provide guided instruction and counseling that encourage interns to reflect on their teaching and its effects on learners, assist in the acquisition of wider repertoires of teaching strategies, and relate problems of teaching practice to research on teaching and human development. The internship should also provide opportunities for interns to prepare for professional practice by participating in institutional and departmental decision making and reviewing policy and practice.

If internship programs in teaching were modeled after those in other professions, they would differ from current beginning-teacher evaluation programs in a number of ways:

- The intern would not only work directly with clients but would also observe and learn from experienced professionals' interactions with clients. Learning by modeling is a precept of an internship. Learning only by doing — that is by trial and error alone — is what beginning teachers do now, and it is much less effective.
- The intern would assume progressive degrees of responsibility for client services.
- The intern would receive regular supervision and guidance from practicing professionals.
- The programs would require that all interns experience *particular* types of situations for decision making and practice under supervision, including a range of tasks and types of clients.

Unfortunately, many of the current state-mandated beginning-teacher programs focus so much on evaluation that they have neglected the supervision and guidance aspects of induction. Because these programs are tied directly to licensure decisions, they concentrate on training candidates to perform to the specifications of particular evaluation instruments. However, they do little to encourage candidates to teach reflectively; to evaluate what they are doing and assess whether it's working and why; to understand how to make better decisions; or to learn how to juggle the various concerns of teaching.

Fox and Singletary (1986) observe that few of these new beginning-teacher programs focus on providing the novice with the assistance required to ease the critical transition from student to teacher. The authors point out that "few (programs) focus on the goals of developing a reflective orientation and the skills essential to self-evaluation." Without these essentials, beginning teachers are likely to be frustrated and unable to cope with the demands placed on them. In fact, the authors have concluded that beginning-teacher evaluation programs devoid of structured assistance increase rather than reduce pressure on new teachers.

Borko (1986) observes that because the programs were created primarily to evaluate candidates for state certification, they fail to provide useful training:

> Most state-mandated (beginning-teacher) programs require that beginning teachers demonstrate competence in a standardized set of teaching behaviors in order to receive certification. Assistance is often viewed as remediation and limited to observed deficiencies in the generic teaching competencies assessed within the program. Because certification criteria must be consistent across the state, most programs are not context-responsive. Moreover, the primary function of state programs is gatekeeping or screening. Thus, competing concerns for individual teachers' professional growth on the one hand, and for establishing a defensible data base to support a recommendation against certification on the other, often shape the nature of assistance.

If teaching were to follow the lead of other professions, beginning teachers would not be evaluated for licensure on the basis of on-the-job observations. Instead, the internship would be a distinctly separate step en route to the licensing exam. In states that have tried to combine

induction with licensure decisions, the practice has suffered from three major shortcomings:

- The rating instruments have failed to take teaching context or content into account because, in an attempt to be objective, they have specified a single set of uniform teaching behaviors to be tallied in a small number of classroom observations.
- The assessment systems have failed to guarantee reliable assessment across candidates because they have evaluated candidates in diverse job settings and performance situations, thereby compromising the fairness of licensing decisions.
- Licensing assessments have been made in part by employers who were also responsible for hiring and granting tenure, thereby entangling licensing and employment decisions (Wise and Darling-Hammond, 1987).

The fact that no other profession now evaluates on-the-job performance as the basis for state licensing should suggest that on-the-job evaluation may be problematic. On reflection, the reasons become clear. Teachers are not licensed to instruct a particular group of children, such as "fifth graders at Kennedy Elementary." Instead, they are licensed to teach children who differ with respect to grade level, general intellectual ability, academic achievement, stages of cognitive development, educational opportunity, socioeconomic status, family attitudes toward education, and many other characteristics. Assessing a candidate's ability to instruct children in one classroom provides little information about whether that candidate is likely to be effective in teaching children with very different characteristics and educational needs. Furthermore, how well a teacher performs on the job is a function not only of his or her knowledge, skills, and dispositions but also of the teaching environment. It is, in principle, unfair to assess a teacher's performance without taking these factors into account. Otherwise, random factors — district resources, curricular approaches, student characteristics, and so on — will determine how well a person appears to perform.

All interns should experience particular types of situations for decision making and practice, including a range of tasks and interactions with different types of students. In taking clinical training seriously, we need to identify those things that all beginning teachers should encounter and learn to master, rather than allowing happenstance to

determine them. Since teachers are licensed to teach *all* students in the subject area(s) or grade levels for which they receive that license, they should be taught how to work with a range of students. The profession should specify a set of educational goals for the internship experience and incorporate them into the design of programs. If teaching were to follow the lead of other professions, candidates would complete that internship before sitting for the clinical practice examination.

Benefits of an Internship

The development of residency or internship programs in teaching will serve some other goals as well. Rates of attrition from teaching are very high in the early years: Most studies find that 30 to 50 percent of new teachers leave within the first three to five years (Grissmer and Kirby, 1987). During a period of teacher shortage, it is especially important that the profession not lose one-third to one-half of the people who enter. A setting that provides assistance and support to beginning teachers can make a tremendous difference in retention. One of the greatest predictors of committing to the profession is a sense of efficacy, the sense a teacher has that he or she is making a positive difference in the lives of students (Rosenholtz and Smylie, 1983; Bredeson, et al, 1983; Chapman and Hutcheson, 1982). People do not go into teaching for the money, but for the satisfaction of making that contribution. Many teachers leave if they feel inadequately prepared for the job and consequently unsuccessful in reaching students in the way that they would like. Making beginning teachers more successful will enhance commitment to the occupation.

Separating the formative assistance task in a residency from the summative examination should encourage reflective teaching. It is very difficult to be openly reflective and critical about one's work if one is also being evaluated for licensure at the same time and in the same context. This two-pronged approach allows the residency or apprenticeship program to be an honest, evaluative learning experience that enhances and encourages reflective practice.

In addition, it is in the internship setting that one can encourage the development of ethical standards and assess, in the case of teaching, whether the person has the kind of concern and care for children that is necessary to become an honorable member of the profession. While professions acknowledge and seek to develop ethical and moral com-

mitment as very important aspects of professional practice, these aspects cannot be assessed well in a paper-and-pencil test. Just as there are some kinds of learning that can only be developed in an internship program, there are also some things that can only be evaluated in the context of actual practice.

The supervisor's role includes ensuring that interns receive thorough clinical preparation, giving frequent and formative evaluative feedback, and explicitly fostering *reflection on practice* and professional ethical awareness. For example, psychology internships specifically seek to instill — along with technical skill — an understanding of professional and ethical responsibility and an appreciation of the intern's own personality, biases, strengths, and weaknesses. Similarly in medicine, the internship and residency requirements stress the candidate's ability to reflect on his or her practices and to demonstrate a devotion to ethical standards. Architecture, too, explicitly seeks a quest for further learning as an outcome of the internship.

These attributes cannot be fully developed in formal preparation programs or fully tested in examinations. But they are legitimate, indeed extremely important, bases for professional licensure. The internship does allow a meaningful setting for developing, observing, and verifying these kinds of abilities.

The internship may provide occasions and events for assessment as well. For example, in some medical-specialty board examinations, doctors must evaluate specific cases they handled during their residency and explain why they did what they did. In architecture and engineering, a portfolio is built up over the course of the residency. Using the internship experiences as one foundation for testing can heighten the validity of the assessment by using real-world occasions as one basis for evaluating practice.

But the internship should not be viewed only as a prelude to testing. As Lee Shulman pointed out this morning, the actual teaching experience is a more complete and valid representation of teaching performance than any examination — written or performance-oriented — can be. The internship is the only setting in which certain professional goals, such as inculcating and evaluating the moral and ethical components of teaching, can be pursued in a safe manner. It is a setting in which context and history can be taken into account, an occasion for structured teaching of clinical skills, and an opportunity for valid assessment of professional capacities that go beyond testable knowledge.

In David Berliner's terms, the internship is the first step on the path from novice to expert status. It is the place where interns can develop strategic knowledge about when general rules apply and when they don't. Here they can develop the kind of proficiency that requires practice; they can develop that sixth sense of appropriateness in judgment and timing; they can make sense of what is observed and experienced; and they can come to know what to pay attention to, what matters, and what has instructional significance.

The internship should pursue both modeling and mentoring. Interns should be able to observe and mimic what experts do, even when it is not articulated, as well as profit from the tutelage of someone who can analyze teaching acts and communicate that analysis. The internship should offer opportunities for observation, practice, debriefing, counseling, and consultation, along with seminars and other didactic experiences. Many of the features that Judy Lanier listed as things to look for in an internship are encapsulated in the apprenticeships of other professions. The combination of those experiences with valid assessment of beginning skills reassures the public that all entrants to the profession have mastered the basic knowledge and skills needed to perform responsibly before the entrants are licensed to practice independently. In addition, this approach holds promise for improving the overall quality of teaching practice by acknowledging the complexity of teaching and encouraging the acquisition of a broad set of understandings and abilities, rather than a formulaic set of behaviors that is not ultimately useful.

Until now, most states have relied either on multiple-choice paper-and-pencil tests of professional knowledge or on-the-job assessments of performance for beginning teachers. The former approach cannot adequately assess the ability of teacher candidates to apply knowledge with sound judgment in complicated, nonroutine situations. The latter approach, unfortunately, does not provide a reliable and generalizable assessment of teaching knowledge and abilities. The benefits of a new approach, one that balances the profession's need for supervised clinical training as well as assessment, and the need of a licensure process for validity as well as reliability, will be many. And foremost among them is the development of public confidence in a competent teaching profession.

References

Borko, H. (1986). Clinical teacher education: The induction years. In J. V. Hoffman & J. Edwards (Eds.), *Reality and reform in teacher education*. New York, NY: Random House.

Bredeson, P. V., Fruth M. J., & Kasten, K. L. (1983). Organizational incentives and secondary school teaching. *Journal of Research and Development in Education, 16*, 52-56.

Chapman, D. W., & Hutcheson, S. M. (1982). Attrition from teaching careers: A discriminant analysis. *American Educational Research Journal, 19*, 93-105.

Darling-Hammond, L., Wise, A. E., & Pease, S. (Fall 1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research, 53(3)*, 285-328.

Fox, S. M., & Singletary, T. J. (1986). Deductions about support induction. *Journal of Teacher Education, 37*, 12-15.

McDonald, F. (1980). The problems of beginning teachers: A crisis in training. *Study of induction programs for beginning teachers* (Vol.1). Princeton, NJ: Educational Testing Service.

National Institute of Education. (1979). *Beginning teachers and internship programs*. Report Number 78-0014. Washington, D.C.: U.S. Department of Education.

Rosenholtz, S. J., & Smylie, M. A. (December 1983). *Teacher compensation and career ladders: Policy implications from research*. Paper commissioned by the Tennessee General Assembly's Select Committee on Education.

Ryan, K. (1979). Toward understanding the problem: At the threshold of the profession. In K. Howey and R. Bents (Eds.), *Toward meeting the needs of beginning teachers*. Minneapolis, MN: U.S. Department of Education/ Teacher Corp.

Tisher, R. (Ed.) (1978). *The induction of beginning teachers in Australia*. Melbourne, Australia: Monash University.

Wise, A. E., Darling-Hammond, L., & Berry, B. (January 1987). *Effective teacher selection: From recruitment to retention*. Santa Monica, CA: The RAND Corporation.

Wise, A. E., & Darling-Hammond, L. (November 1987). *Licensing teachers: Design for a teaching profession*. Santa Monica, CA: The RAND Corporation.

Will Improving Teacher Assessment Improve the Education of Children?

P. Michael Timpane
Columbia University

I have been to more than a few conferences like this and so have you; you should know and appreciate that you have heard the best ever from the best today. I was loathe to prepare my remarks before I heard what the speakers preceding me had to say — but I went ahead anyhow. You have heard from a profusion of eminence about subtle new motions and examples of progress in this area of teacher assessment. And yet I am asked to address the toughest question of all — the bottom line.

Greg Anrig asked me to "bring the conference to a thoughtful conclusion" and then answer the question, "Will teacher assessment improve the education of children?" I am less expert in this field than many of the people who have spoken, and perhaps I am more skeptical. It is not in my nature to be downbeat, but this is the hardest question of the day to answer "yes" to. My remarks elucidate, I believe, what has been a subtext of all the speakers, but it is not one they could emphasize very much because of the nature of their particular assignments.

We can have no doubt that the movement toward teacher assessment in this nation is substantial, durable, and nationwide. Whether we are talking about admission to teacher-preparation programs, certification exams, assessment as part of the induction or the professional-licensure process, or the ongoing evaluation of schools, the role of assessment grows steadily in each of these dimensions. As Larry Rudner has suggested in the past few years, this is part of a long-term movement, a quest, to locate accountability in our schools. For many years, we thought the testing of students would establish this accountability, and when that didn't work, we tried the evaluation of programs, and when that didn't work, we began to zero in on the perfor-

mance of the individual educator. The focus changes from time to time, but the quest is not new. And it is a growing feature of our society, not simply of our educational systems.

It is a logical progression in a sense, but it is far too often associated (as has already been remarked on today) with the quest by policymakers for control of practice rather than simply for accountability — or, more dangerously, with a quest by policymakers for a magic potion to restore public confidence in education. In this case, assessment is merely a misplaced hope. But there is nothing in the policy world that I know of that suggests that this direction toward greater accountability and more substantial teacher assessment is going to change dramatically. The pace may change; the details may change. Some ideas and directions may prove to be worthwhile, some not. But we will not, as a nation, retreat from this quest. It is a part of the larger *quid pro quo* that seems to govern the educational reform movement today — that if there are going to be more resources for the schools and better pay — the *quid* — then the policymakers of the world are going to insist on a *quo*. They are going to be looking at outcomes, somehow defined. And that, I believe, is the nature of the deal in their minds. We may have problems with that, we may wish for a different world, but those are the only terms we have, for the present at least.

My method of preparing these remarks to try to answer the question, "Will teacher assessment improve the education of children?" has been rather informal. I have conferred with people more knowledgeable than I, whose judgment I respect, on the faculty at Teachers College and elsewhere, including some who are in this room today. From that review I have distilled four answers to my question. They are, first, "No, it won't matter;" second, "No, it will hurt;" third, "Compared to what?" and fourth, "It all depends."

Those who respond, "No, it won't matter," believe that teacher assessment will be too weak an instrument, both technically and instrumentally within the world of education, for the intended purpose. A cynic I spoke with put it as follows: "At best, teacher assessment is like a Wasserman test. If you pass it, it doesn't mean much. If you fail, you're in big trouble." Hyperbole aside, it seems to me that this point of view says that there is no strong reason today, in knowledge or experience, to think that teacher assessment will make a substantial difference in student improvement. It takes the conservative, scientific position that the burden of proof should therefore be on those who say it will.

The second school of opinion, "No, it will hurt," suggests that teacher assessment will inevitably, or at least most probably, magnify the technical and routine aspects of teaching and of the teaching role, and will similarly overemphasize the more readily measurable aspects of teaching — commit the empirical fallacy, in other words. Some of the concerns we heard from David Berliner this morning about intuition and arational expertise as the hallmarks of various levels of successful teaching suggest how tempting that fallacy will be and how difficult it will be to measure these characteristics. This school of thought takes some comfort, probably out of context, from the sayings of Greg Anrig, who has said from time to time, "Tests don't measure dedication, caring, perseverance, sensitivity, and integrity." This school suggests that tests never will, and that they will therefore measure other things, to the detriment of education. It also suggests what has been suggested by many today, that teacher assessment (at least as it is currently being constructed) will make it more difficult for less-well-prepared test takers to enter and progress in the profession.

The third school of thought is the "Compared to what?" school. It simply suggests that other policies and other decisions will be far more determinative and far more important for improved student performance, and that teacher assessment should be merely in the service of these policies. This school of thought argues, on the positive side, that more money and sustained political support will certainly make more difference than better teacher assessment in improving the education of children. Better pay and status for teachers, widespread success with school-improvement strategies, reforms in teacher education, and research findings that unlock some of the secrets of learning — these are the matters that will determine whether or not the education of children improves. On the negative side, this school asks the question, "Can better teacher assessment possibly do anything about overregulation, overstandardization, and top-down controls in curriculum and instructional methods, or chronically poor morale, burnout, and inadequate training and professional development?" The answer, of course, is "no."

The fourth view, and the most helpful and realistic, is "It all depends." In this view, the "no" answers are problems to be surmounted and pitfalls to be avoided. The "Compared to what?" answers provide an appropriate policy context in which teacher assessment can flourish or fail, as well as a measure of modest, realistic expectations for teacher assessment per se.

It all depends on many things. It all "depends," for example, on whether or not the assessment is designed and carried out in an organizational and a systems perspective, in league with the many other policies of school improvement, better teacher preparation, improved teacher-induction programs, and many other factors. The critical question becomes, "Is the assessment part of a set of policies trying to move teachers and schools in the same direction, and is it put to appropriate use within that context?" It also depends on whether the assessment has an appropriate conception of teaching — does it think of teaching as an active, heuristic, and dynamic process, or one that is mechanical and governed by a set of rules? It depends on whether the assessment involves teachers in its design, administration, and interpretation. And it depends on whether the assessment is, in the context of the work life of the schools, technically complete: Is it congruent with the assigned responsibilities of the individuals? Does it have clearly identified and agreed-on criteria? Is it based on appropriately collected information? Does it provide accurate and consistent appraisal of that information, and feedback and planning for improvement? It also depends on whether or not the assessment captures the consistent aspects of performance of what one knows or does that make a difference. (You notice I have avoided the "v" word — validity. Validity, in my opinion and experience, causes pandemic narcolepsy at four o'clock in the afternoon. I thought if we all just concentrated on whether the assessment captures the consistent aspects of performance that make a difference, we would have all the precision we need at this time of day.)

In threading through these four, five, or six considerations, and in coming up with all the "whethers" and "nots" with respect to them, we still encounter some dilemmas in trying to reconcile them. Reconciling the policy objectives of professional autonomy and professional accountability, on the one hand, and the necessary cost and complexity of appropriate assessment systems, on the other — it is far from obvious how one threads that path. It is far from obvious how one reconciles collegiality with the assignment of causality and individual responsibility to one particular teacher. Many other dilemmas must also be worked out.

All in all, I conclude that the path to an effective teacher-assessment system that is worth the trouble will be tortuous indeed. The link to students' outcomes will be tenuous and will probably even then be inferred from other phenomena in the development of teacher professionalism and in the reform or restructuring of the schools.

The big question — and I think it is the question that has been before this group all day — is, "What is the [illegible]pe of the role of teacher assessment in the professionalization project we are embarked on in our nation?" "Unavoidably large," is the answer, from all the evidence we have heard today. And troublesome, too. But the fact remains that every profession in our society must define, develop, and operate a professional assessment system and satisfy the public about it. We have not yet done so in any systematic sense in education. We are working busily on the task, more busily and more productively than at any time I can remember, but we have not a moment to spare.

The announcement made yesterday by ETS about the successor to the NTE and the ambitious plans that Jim Kelly described to you this morning regarding the National Board for Professional Teaching Standards suggest that we will have large, enlightened, intelligent, comprehensive projects addressing these issues for the next several years, more so than ever before.

So I say good luck, Greg Anrig and ETS, and good luck, Jim Kelly and the National Board for Professional Teaching Standards. May you, again in Greg Anrig's terms, demonstrate the dedication and the caring and the perseverance and the sensitivity and the integrity that the profession and its assessment must also identify and strengthen.

06357 - 00477 • 539M2 • 272533